2020

Conceptions of Internalizing and Externalizing Symptoms: Identifying Parent and Child Differences Using Psychometric Analyses

Jenna Winarick

CONCEPTIONS OF INTERNALZING AND EXTERNALIZING SYMPTOMS: IDENTIFYING PARENT AND CHILD DIFFERENCES USING PSYCHOMETRIC ANALYSES

A dissertation submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

to the faculty of the department of

PSYCHOLOGY

of

ST. JOHN'S COLLEGE OF LIBERAL ARTS AND SCIENCES

at

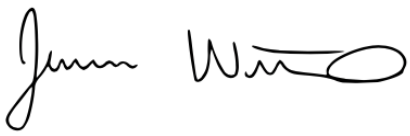
ST. JOHN'S UNIVERSITY

New York

by

Jenna Winarick

Date Submitted __04/03/2020_____ Date Approved___04/03/2020____



Jenna Winarick William Chaplin, Ph.D.

ABSTRACT

CONCEPTIONS OF INTERNALZING AND EXTERNALIZING SYMPTOMS: IDENTIFYING PARENT AND CHILD DIFFERENCES USING PSYCHOMETRIC ANALYSES

Jenna Winarick

One of the most prevalent ways to assess child and adolescent symptomatology is with a multiple-informant approach (Kraemer et al., 2003; De Los Reyes et al., 2015), yet mental health professionals commonly run into the problem of informant discrepancies in ratings of symptomatology (Achenbach, 2006). These comparisons are made on the observed scores; however, we argue that to simply compare mean scores may not be accurate and instead, researchers must examine the role of measurement invariance. We hypothesized that information discrepancies between parent and self-report on the YOQ is due to lack of measurement invariance. Our assessment of measurement invariance was based on differences in the factor structure and differential item functioning of the YOQ between parents and children at two timepoints (N Parents T1=70; N Parents T2=63; N Children T1=79; N Children T2=74) using Internalizing and Externalizing factor structures. The results from our study show that children and parents perceive symptoms differently, especially behavioral indicators of internalizing disorders. We conclude that direct comparisons of parent and child observed scores on the YOQ require the treating clinician to understand and acknowledge these discrepancies when developing an assessment-based case conceptualization and collaborative treatment plan. Implications and future directions are discussed.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my mentor and research advisor, Bill Chaplin, for his continuous support and guidance throughout this process. Bill, you have been an amazing role model in more ways than one. Throughout the past five years you have taught me what it means to be an intelligent creator and consumer of research, and perhaps more importantly, what it means to be an effective leader. Your dedication to your students is bar none and I can only hope to emulate the amount of patience, understanding, and humor you bring to your work each and every day.

I would also like to thank the members of my committee, Alice Pope and Andrea Bergman, for their roles in this project. Alice, I have had the pleasure of your mentorship since my first day at St. John's and it has helped me grow both professionally and personally. You are not afraid to stand up for what you believe in, and you do it with tremendous grace and perseverance. I've never met anyone so willing to do what needs to be done, even when it is difficult or unpopular. You are truly an inspiration. Andrea, I feel incredibly grateful to have had your clinical expertise serve as a guide from the early days of my training. Your depth of knowledge, empathy, and passion for this work has helped me to develop as both a researcher and clinician.

To my husband, thank you for encouraging me to pursue my dreams. Your love and support mean the world to me and I am a better person because of you. You are always there to cheer me on and celebrate all of the milestones along the way, and it has made all of the difference. I love you tremendously.

And lastly, I would like to thank my mother. Mom, you were my first role model and continue to be my biggest inspiration. I would not have been able to accomplish any

of this had it not been for your own sacrifice, patience, and unwavering love. You raised me with an understanding of what it means to work hard and follow my instincts, and for that I am forever grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

The ability to track change over time is imperative to the field of clinical psychology. In recent years, there has been an increased demand for the use of outcome instruments in mental health practices due to a variety of research, clinical, and economic advantages. First, outcome measures provide treatment providers with an assessment of baseline functioning, that is, the current level of symptomatology prior to receiving treatment (Lambert et al., 2001; Wells et al., 1996). Second, when used in a repeated measure design, outcome measures can track change throughout treatment. Clinicians continuously state that they find outcome measures most valuable for tracking clinical progress throughout treatment (Hatfield & Ogles, 2004). Third, clinicians report that they use standardized outcome measures to determine the course of treatment (2004); finally, outcome measures inform a more ethical standard of practice (2004).

Economically, increased use of outcome measures is partly due to standards set by managed healthcare organizations and private insurance companies (McClendon, 2009; Sanchez & Turner, 2003). These organizations were developed to provide appropriate, cost-effective treatment, and are more likely to extend benefits to mental health treatments that demonstrate reliable and significant change (Koss & Shiang, 1994). Moreover, practitioners are increasingly encouraged to provide evidence-based treatments while optimizing the number of sessions to achieve a reduction in symptomatology (McClendon, 2009). Thus, clinical psychologists and other mental healthcare practitioners favor instruments with high sensitivity to change, as well as excellent reliability and validity, to demonstrate treatment effectiveness (Hatfield & Ogles, 2004; McClendon, 2009).

Complications in tracking meaningful variations in clinical outcome data arise when measurements are assumed to have a consistent relation across time between the observed score and the underlying latent outcome (e.g., symptoms, diagnosis, distress) the score represents. In addition, clinicians and researchers typically obtain outcome measures from both children and their parents, and the observed scores on these measures are often discrepant. The primary aim of this research is to consider informant discrepancies and the temporal consistency of measures obtained by those informants over time using psychometric models based on latent-trait theory.

Multi Informant Assessment Discrepancies

In the field of child and adolescent psychology, treatment providers consistently turn to multi-informant assessment measures to provide a more thorough understanding of a client's current symptom presentation (Kraemer et al., 2003; De Los Reyes et al., 2015). The multiple informants generally include the parent and the child but may also include the child's teachers, peers, and the treating clinician. Obtaining these measures is standard practice at mental health clinics and clinicians use them to track information about client wellbeing. However, one of the most common issues that practitioners face is that of informant discrepancies (De Los Reyes & Kazdin, 2005). Multiple informant discrepancies can become problematic when mental health practitioners and insurance companies use that information to track change, determine treatment effectiveness, and decide when to terminate treatment. Beyond that, these discrepancies can create problems such as conflicting treatment recommendations and inconclusive research findings (De Los Reyes, Kundey, & Wang, 2011).

In one study examining discrepancies between parent, child, and therapist ratings of child mental health symptoms, the authors found that more than three-quarters of the triads (child/parent/therapist) began therapy without agreeing on a presenting problem (Hawley & Weisz, 2003). They also highlighted where informants had higher and lower levels of agreement and disagreement. That is, parents and therapists had higher levels of agreement on the child's presenting problems; however, the children and therapists had higher levels of agreement on family and environmental problems (2003). These findings emphasize the possible clinical implications if clinicians use the results of a single informant to inform treatment planning and make recommendations, or if the clinician weighs the report of one informant more so than another.

For instance, De Los Reyes et al. (2011) found that clinical interviewers perceived children as having more behavioral concerns and were more likely to rate them as an unreliable informant when parent reports showed greater levels of behavioral concerns compared to youth self-report. Conversely, these interviewers were less likely to rate the children as unreliable when their parents endorsed lower levels of behavioral concerns (2011). Giving more weight to one respondent over another based on individual perception of specific symptomatology can have detrimental effects on the identification of mental health disorders and subsequent treatment of children and adolescents.

Previous research has posited several hypotheses about the underlying causes of informant discrepancies. These hypotheses theorize that discrepancies exist due to different psychological influences in the informants (i.e., self-perception, social desirability, etc.), different contexts in which the symptoms occur (i.e., school vs. home), symptom type (internalizing vs. externalizing), sociocultural considerations, and

psychometric factors including measurement error. De Los Reyes et al.'s (2015) meta-analysis examining the validity of a multi-informant approach to the assessment of child and adolescent mental health addresses each of these hypotheses in depth.

One possible explanation as to why informant discrepancies occur may be due to factors that influence a child or adolescent's self-perception (Cleridou, Patalay, & Martin, 2017). Self-reports are often seen as observable manifestations of an individual's self-perception; therefore, it may be possible that mental health disorders that impact self-perception will also impact how children and adolescents respond to measures of mental health. For example, research has found that children with Attention-Deficit/Hyperactivity Disorder (ADHD) predominantly hyperactive/impulsive subtype have higher positive illusory self-perceptions of their behavior than controls (Owens & Hoza, 2003). This research suggests that children with predominantly hyperactive and impulsive behavioral problems due to ADHD may be less likely to rate these behaviors as problematic compared to parent and/or teacher-report because they may simply lack the self-awareness.

Another condition that may impact self-perception of mental health symptomatology is depression. One of the mainstay theories of depression is Beck's cognitive triad, which specifically highlights alterations in self-perception as underlying the existence and severity of depressive symptoms. Beck's theory posits that there are three key elements that impact the cognitive processes including automatic negative thoughts about the self, the world or environment, and the future (Beck, 1979). Therefore, it is likely that a child or adolescent suffering from depression may respond differently to items on a self-report questionnaire than an outside observer (i.e.,

parent/teacher) because of the implications of these negative cognitive biases in self-perception.

Social desirability biases may also impact self-perception and ratings of mental health symptoms on self-reports. De Los Reyes and colleagues found that differences between parent and child self-report may be due to variation in the perception of certain types of symptoms, specifically in relation to social desirability biases. For example, researchers found significantly lower levels of self-reported social anxiety among clinic-referred adolescents relative to the parents' reports; however, these self-reports showed minimal correspondence with objective measures of physiological arousal (De Los Reyes, Aldao et al., 2012; De Los Reyes et al., 2015). This relative under-reporting of social anxiety symptomatology in self-report compared to parent-report may reflect an interaction between the anxiety itself and the influence of social desirability factors.

Another explanation is that the differences between parents and child reporting of child distress partly differ as a function of the types of symptoms (De Los Reyes et al., 2015; Achenbach et al. 1987). Specifically, parents tend to be more aware of observable, externalizing behaviors (e.g., aggressiveness, hostility, rule-breaking, hyperactivity) compared to internalizing symptoms (e.g., fears, worries, and restlessness). Achenbach and colleagues (1987) first established this finding in one of the first meta-analyses examining informant discrepancies in child psychological assessment. Achenbach et al. (1987) found informants agreed more on reports of children's externalizing versus internalizing concerns. More recent meta-analyses (De Los Reyes et al., 2015) further bolster these findings, leaving psychological researchers to conclude that informants

differ on their endorsement of symptomatology partly due to the function of the type of symptom.

Additionally, research has investigated whether socio-cultural factors, such as ethnicity, relate to discrepancies in parent and child reports. Van de Looij-Jansen et al. (2010) found evidence for this conclusion in their study investigating parent-child disagreement of internalizing problems. Specifically, they found that children from Surinamese/Antillean backgrounds reported more internalizing problems than their parents, and the reverse was true for children of Turkish descent. However, other studies examining parental culture as a potentially influential factor in parent-child report discrepancies did not find support for this conclusion (Weisz and McCarty, 1999).

In the data collected at our Psychological Services Center, we have found clear discrepancies between parent and child assessment of the child's level of symptomatology as well as the degree to which symptoms change over the course of psychotherapy (Jerusalmi, 2015). Mixed-effects regression analyses revealed that on average, children perceive more change on a weekly basis than do the parents of the children. Although children (over 12) believe that their YOQ symptoms are decreasing .11 per week, parents (of children over 12) only believe their children's YOQ symptoms decrease by .04 per week (2015). These results are in line with the previously cited literature on multiple informant discrepancies and lead to the question: what factors underlie these discrepancies in perceived symptomatology?

Discrepancy in Measures Over Time

Scores on measures typically vary over time within each informant. Score variation is most likely to occur when people put forth effort to change the underlying

construct that the measures assess. This is certainly the case for outcome measures obtained over the course of psychotherapy. However, there may be several reasons that psychological measures are discrepant over time besides a change in the underlying construct. These can include contributing factors such as practice effects, technology (paper-and-pencil vs tablet), developmental effects, inattention, and others.

Practice effects are a problematic characteristic of a repeated measures design attributed to increased familiarity with the assessment that aids in procedural learning (Kelly & Garavan, 2005; Budson & Price, 2005) and strategy development to help group information more effectively (Goldberg et al., 2015). Researchers have found that practice effects influence the outcome of neurocognitive assessments when people take the same test several times within a short time period (Duff et al., 2007; Bartels et al., 2010; Goldberg et al., 2015). For instance, Duff et al. found that practice effects accounted for as much as 83% of the variance in follow-up tests of cognition in adult participants (2007). Practice effects can certainly influence how parents and children respond to items on psychological outcome measures given the frequency of repeated administration. Individuals may learn the item grouping or scoring system over time and inadvertently allow the way in which they processed the items on previous administrations to influence their responses on follow-up measures.

The test medium may also contribute to discrepancies in measures over time within the same respondent. Many mental health facilities now use electronic measures administered on tablets such as iPads and Kindles. Technology allows for benefits such as more efficient testing, accessibility, and immediate feedback. There have been several studies that demonstrate the high correlation between paper-and-pencil versus electronic

forms of test administration (Wolford et al., 2008; Cook et al., 2007; Gwaltney, Shields, & Shiffman, 2008; Goldstien et al., 2010; Joubert & Kriek, 2009).

However, participants consistently rate technological forms of administration as more preferable to paper-and-pencil. A meta-analysis of 30 publications that included preference data found that 87% of the participants preferred electronic to paper-based forms of assessment (Campbell et al., 2015). These findings may be particularly applicable to children and adolescents due to the prevalence of technology in their everyday life. Therefore, individuals may be more likely to complete a measure to the best of their ability if it is provided in electronic, compared to paper-and-pencil, format given the preference for this assessment medium. The impact of this research is especially relevant if the assessment medium has changed over time within the same respondent group. For example, if researchers administered the outcome assessment for one person in paper-and-pencil format and now administer the assessment on a tablet.

Developmental effects, such as hormonal changes and cognitive/social maturation, may also impact how individuals respond to psychological assessment measures over time. Developmental effects can be especially pertinent to adolescents who are undergoing significant developmental changes on a daily basis. Hormonal changes, for instance, can affect levels of restlessness and fatigue (Peper & Dahl, 2013), which may significantly impact how individuals respond to an assessment measure at one timepoint compared to the next.

Developmental effects also play a role in the psychotherapeutic paradigm itself, such that that psychotherapy intends to increase the level of understanding of specific symptoms and therapeutic processes over time. Throughout the course of many

evidence-based psychotherapies, clinicians routinely engage the client (and caregiver when applicable) in psychoeducation of the presenting concern (Lukens & Mcfarlane, 2004). Psychoeducation is particularly important in the first few months of therapy because it provides information, helps clients and their families to set goals, clarifies questions and concerns, and builds commitment to treatment (Hayes, Harvey, & Farhall, 2013). Hence, it is possible that these processes will impact the understanding of symptomatology assessed using a psychological outcome instrument, and thus have an effect on measurement discrepancy over time.

Lastly, an individual's level of attention may vary significantly at one timepoint compared to another. As attentional levels wax and wane, so may responses on psychological questionnaires. For example, when a client has gotten a restful night's sleep, eaten three meals, and is not cognitively overloaded with pressing issues or concerns, he or she may have an easier time attending to a psychological assessment measure than they otherwise would if they experienced factors that diminished their attentional capacity.

Taken together, practice effects, technology, developmental effects, and inattention can all influence longitudinal measurement invariance within the same respondent. Given that these factors may influence scores on psychological outcome assessments separately from actual symptom change, it calls into question what we mean when we talk about "change" in each respondent over time. Simply put, "if the pretest and the posttest are measuring different dimensions, no amount of statistical manipulations will produce a measure of gain or growth" (Lord, 1958).

Measurement Invariance

Thankfully, researchers have begun to consider the psychometric properties of child and parent psychological assessment measures as a potential contributing factor to multiple informant discrepancies. Results have yielded several important findings; namely, that measurement error is related to rater biases, and correspondence between multiple informants is highest when (a) the reports assess more observable behaviors (e.g., externalizing as opposed to internalizing), (b) these behaviors occur in the same context (e.g., home vs. school), and (c) they use a dimensional, rather than categorical, approach (De Los Reyes et al., 2015). Thus, psychometric analyses have vital implications for the clinical effectiveness and research validity of multi-informant assessments (Dirks, Boyle, & Georgiades, 2011).

Taken together, these findings suggest the main factors that contribute to discrepancies between parents and children's assessment of symptomatology (De Los Reyes et al., 2015; Achenbach, 2006; De Los Reyes & Kazdin, 2004, 2005, 2006; Goodman, De Los Reyes, & Bradshaw 2010). However, one of the problems in the general literature comparing parents and children is that researchers make comparisons of the observed scores based on the assumption that the same observed score means the same thing about the underlying symptomology whether it is provided by a parent or a child (Hawley & Weisz, 2003). That assumption is not necessarily tenable. Indeed, we argue that to simply compare mean scores may not be accurate; instead, researchers must examine the role of measurement invariance; that is, do responses of children and parents map on to the underlying construct the same way? And does this mapping change over

time? Reise, Widaman, and Pugh (1993) best explain the significance of this research question:

> "To compare groups of individuals with regard to their level on a trait, or to investigate whether trait-level scores have differential correlates across groups, one must assume that the numerical values under consideration are on the same measurement scale. That is, one must assume that the test has 'measurement invariance' across groups. If trait scores are not comparable across groups, then differences between groups in mean levels or in the patter of correlations of the test with external variables are potentially artifactual and may be substantively misleading." (p.552)

There is minimal research on the role of measurement invariance and differential item functioning in multiple informant discrepancies of longitudinal outcome research. In the field of psychology, researchers and practitioners interpret the observed sum score of an assessment measure using the assumptions of classical test theory. These assumptions include that the measure is unidimensional (i.e., reflect a single underlying construct) and that response patterns are monotonic, that is, individuals base their response to all items on the same underlying scale that has a monotonic relation to the underlying construct. We argue that these assumptions are not necessarily valid in the interpretation of outcome measures and progress monitoring.

One study examining four different rating scales of depression highlights the problem of using classical test theory to interpret depression severity scores (Fried et al.,

2016). Results from this study show that all four of the measurements did not meet the standards set by classical test theory, and thus, observed sum scores are not necessarily true markers of depression (2016). More specifically, exploratory and confirmatory factor analyses challenged the assumptions of unidimensionality and specified that scales required multiple factors to accurately capture the underlying construct. Furthermore, longitudinal analyses found that the depression measures violated the assumption of temporal invariance in that the factor structures changed over time (2016). The systematic challenging of classical test theory assumptions is rare in psychological research, and these findings suggest the usefulness in revisiting methods of measurement interpretation.

Current Study

The purpose of this study was to undertake an evaluation of the temporal measurement invariance and differential item functioning of the Youth Outcome Questionnaire (YOQ; Burlingame et al., 2001) between children and parents of children over the age of 11. The Youth Outcome Questionnaire was developed to measure psychological symptomatology over time and detect changes that demonstrate improvement or worsening of symptoms (2001). The YOQ uses a multi-informant (self and parent-report) approach to measure symptom change.

Our assessment of temporal measurement invariance was based on differences in the factor structure of the YOQ between parents and children at two timepoints (Time 1 and Time 2). Although the YOQ has been conceptualized as a single unidimensional scale (YOQ; Wells, Burlingame, & Rose, 2003; Dunn et al., 2005), it is clear that the items can be grouped into several dimensions. A very common categorization of items is

into the broad categories of "internalizing symptoms" and "externalizing symptoms" (Achenbach, 1991), and we hypothesized that we could apply this two-factor structure to many of the items on the YOQ.

This framework for understanding child and adolescent psychopathology, first established by Achenbach (1991), defines internalizing symptoms as those related to anxiety/depression, somatic complaints, and patterns of withdrawal; whereas externalizing symptoms refer to those characterized by "delinquent and aggressive behavior" (Achenbach, 1991; Levesque, 2011). Based on the efficiency and widespread tendency to summarize many symptoms into the broad categories of internalizing or externalizing symptomatology, we will use this two-factor structure of the YOQ as the starting point for our evaluation of measurement invariance.

Previously, we established this two-factor structure (Winarick, 2018) and identified a common set of items between parents and children to assess internalizing (INT) and externalizing (EXT) symptomatology (see Table 3). Specifically, our exploratory factor analyses revealed that there were important differences between the parents and children on the overall YOQ structure. First, when comparing the child data to the parent data, we found that several items switched factors. That is, they loaded on one factor for the analysis using the YOQ-Self Report (YOQ-SR) data, and the opposite factor for the YOQ-Parent Report (YOQ-PR) data analysis. Interestingly, the only items that switched factors all did so in the same direction: from the child INT factor to the parent EXT factor. The four items are "I have a hard time finishing assignments or I do them carelessly," "My emotions are strong and change quickly," "I complain about or

question rules, expectations, or responsibilities," and "I have a hard time concentrating, thinking clearly, or sticking to tasks."

Second, we found that there were items that loaded significantly onto the child INT factor, but the same items did not load onto either factor when we conducted the factor analyses using the YOQ-PR data. These items were: "I don't participate in activities that used to be fun" and "I feel irritated." Third, we found that there were items that did not load significantly onto either factor in both the child and parent analyses. These items were: "I have a hard time sitting still (or I have too much energy)," "I use alcohol or drugs," and "I have threatened to, or have run away from home." Thus, we did not include these items in the confirmatory factor analysis to assess measurement invariance analyses. Finally, we found that there was one item ("I act without thinking and don't worry about what will happen") that did not load onto child INT or EXT factors but did load on the parent EXT factor.

Based on the common set of items, we then undertook a series of confirmatory factor analyses to assess measurement invariance on these item sets. The results of our configural and metric invariance tests reflected that, if unconstrained, the items did not load similarly on the INT factor when comparing parents versus children on the YOQ. Furthermore, these analyses also failed to establish measurement invariance at the least restrained model for the EXT item set. Our results clearly suggested that there are differences between how parents and children respond to YOQ items in relation to the underlying latent variables of INT and EXT symptomatology. Thus, direct comparisons of parent and child observed scores on the YOQ and the internalizing and externalizing sub-factors may not be warranted.

Research Questions

We used these findings from our previous analyses to develop our current hypotheses, which include:

1. Child vs. Parent Comparisons at Time 1
    a. Exploratory factor analyses conducted on the YOQ-SR and YOQ-PR reports at Time 1 will largely replicate the results found in Winarick, 2018. That is, most of the same items will fall on the same factors for both child and parent reports. We also hypothesize that several items will switch factors, and do so in the same direction, from the child INT scale to the parent EXT scale. Lastly, we hypothesize that parents will have more items on their EXT scale and children will have more items on their INT scale at Time 1.
    b. We hypothesize that the two-factor structure will fit better for children than parents at Time 1 as evidenced by a lower eigenvalue ratio of the first and second factors.
    c. We hypothesize that there will be significant differential item functioning as determined by IRT analyses between YOQ-SR and YOQ-PR at Time 1.
2. Child vs. Parent Comparisons at Time 2
    a. We hypothesize that exploratory factor analyses will find more agreement between parents and children as to which items fall on the INT and EXT scales than they did at Time 1 because of the psychoeducation of both parents and children that occurs during

psychotherapy. We also hypothesize that several items will switch factors, and do so in the same direction, from the child INT scale to the parent EXT scale. Lastly, we hypothesize that parents will have more items on their EXT scale and children will have more items on their INT scale at Time 2; however, the discrepancy in the number of items that differ will be smaller than at Time 1.

b. We hypothesize that the comparison of fit of the two-factor structure as determined by the eigenvalue ratio will remain the same as it was at Time 1. That is, that the two-factor structure will fit better for children compared to parents.

c. We hypothesize that there will be less differential item functioning as determined by IRT analyses between YOQ-SR and YOQ-PR analyses than at Time 1.

3. Child Time 1 vs. Child Time 2 Comparisons

    a. We hypothesize that exploratory factor analyses will find that the same items loaded onto the same factors across time points.

    b. We hypothesize that the fit of the two-factor structure as determined by the eigenvalue ratio will remain similar across time points.

    c. We hypothesize that there will be no evidence of differential item functioning as determined by IRT analyses when comparing YOQ-SR at Time 1 and Time 2.

4. Parent Time 1 vs. Parent Time 2 Comparisons

a. We hypothesize that exploratory factor analyses will find more items on the INT factor and less on the EXT factor at Time 2 compared to Time 1 as a result of increased psychoeducation about internalizing symptomatology that generally occurs in the first few collateral or parenting sessions of evidence-based treatment for children and adolescents. Also, there will be fewer items that fail to load significantly onto either factor at Time 2 than there were at Time 1.
b. We hypothesize that the fit of the two-factor structure will be better for parents at Time 2 than it was at Time 1 indicating greater discernment between internalizing and externalizing symptomatology.
c. We hypothesize that there will be significant differential item functioning as determined by IRT analyses between YOQ-PR Time 1 and Time 2.

Method

Participants

This study used archival data collected from a community clinic sample in Queens, New York. Time 1 (T1) is defined as the first recorded YOQ data of children over the age of 11 (N=79; mean age=14.58) and parents of children over the age of 11 (N=70; mean age of children=14.70) within 0-4 weeks of their first recorded appointment. Time 2 (T2) is defined as YOQ data of children (N=74; mean age=14.15) and parents of children over the age of 11 (N=63; mean age of children=13.83) 8-12 weeks into treatment (Table 1). Cases with more than 20% of missing data were deleted from the sample and we computed missing item responses using expectation-maximization algorithms for remaining data.

All participants used in these analyses had consented to both receive treatment and to allow their responses to be included in the clinic research data base.

Instruments

The Youth Outcome Questionnaire Self and Parent-Report 30.2 (YOQ®SR/PR) (YOQ; Burlingame et al., 2001) is a 30-item, full scale questionnaire constructed to assess presenting problems and track change over time. The YOQ®-SR 30.2 measures youth and adolescents' (ages 12-17) perception of their own symptom frequency over the previous seven days, whereas the YOQ®-PR 30.2 measures caregiver perception of their child's symptomatology over the previous seven days. Each of the 30 items is answered using a 5-point response scale ranging from "Never or Almost Never" to "Almost Always." The items are the identical for both the YOQ-SR and YOQ-PR versions; with the exception that the self-report uses "I" (e.g., "I destroy property on purpose") and the

parent-report uses the phrase “my child” (e.g., “My child destroys property on purpose”). The YOQ 30.2 was “constructed to be brief, sensitive to change over short periods of time, and available at a nominal cost while maintaining high psychometric standards of reliability and validity” (Burlingame et al., 2004, p. 238).

Procedure

Clients completed the YOQ-SR/PR 30.2 either via paper-and-pencil format or on a tablet upon arrival at the community-based mental health clinic. The instrument was administered by the front desk as part of a series of questionnaires that the clients were instructed to complete before meeting with their therapist as well as continuously throughout treatment. All pencil and paper data were double entered and reconciled before being entered in the database. Tablet data was transferred electronically into the database.

Statistical Analyses

Exploratory Factor Analyses

Our analyses were informed by prior exploratory factor analyses (EFA) of the YOQ-SR and YOQ-PR at Time 1. These EFA findings confirmed our a priori hypothesized structure of two correlated factors whose content reflects internalizing and externalizing symptoms. The EFA allowed us to identify a set of common items on each factor for the parent and child reports. We then conducted a similar EFA on the current dataset of YOQ-SR and YOQ-PR data at Time 2 to determine: 1) if the common items for each factor remained invariant across time, and 2) if the common items for each factor remained invariant across respondent. We used principle axis factor analysis using the squared multiple correlations obtained by regressing each item on the others as our initial

communality estimates for parent and child comparisons at Time 1 and Time 2. We rotated the first two factors using an oblique rotation and determined if the Internalizing and Externalizing scale interpretation to be sensible in each sample.

Confirmatory Factor Analyses

Additionally, there were enough items in common on each factor in the parent and child samples to form the basis of a confirmatory factor analysis (CFA). We used the CFA to verify the fit of the structure in each sample and to subsequently allow the assessment of configural, metric, and scalar invariance in a two-group CFA comparing Children vs. Parents at Time 1, Children vs. Parents at Time 2, Children Time 1 vs. Children Time 2, and Parents Time 1 vs. Parents Time 2.

Configural invariance is the least restrictive form of measurement invariance as it simply requires that the same items load on the same factors in each group, but that the loadings are not necessarily the same. Configural invariance can be thought of as the baseline against which more restrictive invariance models are evaluated. A more stringent assessment of measurement invariance is metric invariance where the factor loadings are fixed to be equal for the two groups. Metric invariance is established if the decrement in model fit (tested by comparing the model fit chi-squares) is not significant when the factor loadings are fixed. The most stringent form of measurement invariance that we tested was scalar invariance. Scalar invariance is established when both the factor loadings and the intercepts of the items are equal across groups. That is, scalar invariance focuses on both the covariance structure and the mean structure of the items. Scalar invariance is particularly important when the means of the groups on the scales (i.e., Internalizing and Externalizing) are being compared. We used chi-square difference

tests between increasingly restrictive models to evaluate the different levels of measurement invariance.

Item Response Theory Analyses

Item response theory (IRT) challenges the assumptions of classical test theory by taking both the person and item attributes into account when interpreting assessment data. Item response theory estimates item parameters using item characteristic curves (ICC) which identify the relationship between the probability of endorsing a response and the underlying latent trait (i.e., internalizing/externalizing symptomatology) otherwise known as *theta*.

For a fully constrained IRT model, the first step is to ensure that all items have to have the same number of responses. For example, one cannot fit a two-group IRT model if the dataset has a participant who endorsed a 4 in response to a YOQ item T1 and no endorsement of a 4 on the same item at Time 2. To address this problem, we converted that extreme responses in one group that were not present in another group to the next most extreme response. A summary of all modified responses for comparisons across person and time are included in Table 2.

We used the Samejima's Graded Response (SGR) model of IRT to specify a 2-Parameter Item Characteristic Curve the discrimination (α) and difficulty (β) parameter. We analyzed the data using simultaneous estimation of both groups and constrained item parameters to be equal across groups (parent/child, T1/T2). Differential item functioning is said to occur when the ICCs of two groups (parent and child; Time 1 and Time 2) differ on one or both parameters. There are many ways to detect DIF including Lord's chi-square, Raju's area methods, and differential test functioning; however, we used the

likelihood ratio test to detect DIF within our sample. One advantage to using the likelihood ratio test is that it can handle polytomous data, such as that provided by the YOQ. Another advantage is that it offers a direct test of non-uniform DIF which occurs when an item is more strongly related to the underlying latent trait in one group than the other for individuals with the same level of theta (general psychopathology).

We tested DIF by following a series of steps outlined by Tay et al. (in press). First, we produced a model in which all items were tested for DIF and examined the p-values (p<.05=significant) that tested the difference between reference and focal group item parameters. We then selected items that did not have significant DIF as "anchor items" and entered all items as "candidate items" for the next model. Next, we tested whether non-anchor items had significant DIF (p<.05) by following the same steps and continued this iterative procedure until all non-anchor items were found to have significant DIF.

Results

Children vs. Parents at Time 1

Initial Tests of Difference

T-Tests. We first compared the initial observed scores of parents and children on the YOQ Total (27 items), Internalizing (INT) subscale (7 items), and Externalizing (EXT) subscale (7 items) in data collected at our psychological service center. The INT and EXT subscale items used were only the items that remained stable across both respondent and time. In contrast to literature that suggests that parents and children often provide ratings that are discrepant, we found similarity between parent and child reports on the YOQ total, INT, and EXT scales at Time 1. We did not find a significant difference in the scores for children (M=29.85, SD=16.77) and parents (M=29.43, SD=18.32); $t(147)=-0.15$, $p=.88$ when comparing the YOQ Total observed score. These results suggest that children and parents of children over the age of 11 do not differ on their initial observed reports of clinical distress.

Furthermore, we did not find a significant difference for children (M=8.63, SD=6.25) and parents (M=7.60, SD=5.33); $t(147)=-1.08$, $p= .28$ when comparing scores on the INT subscale. The same was true for the EXT subscale in which we also did not find a significant difference between children (M=6.59, SD=4.71) and parents (M=6.61, SD=5.55); $t(147)=.02$, $p=.98$. However, to simply compare these mean scores between parents and children requires established measurement invariance. This research was conducted to address this issue and to determine if the lack of difference in the observed scores is a valid reflection of a lack of difference in the underlying latent variables.

Correlations. Correlations between both children and parent's scores on the YOQ Total and INT subscale, YOQ Total and EXT subscale, and INT and EXT subscales were all significant (Table 5). We found that the INT subscale for the children was more highly correlated with the YOQ Total than the EXT subscale. For the parents, we found that the correlations of the EXT and INT subscales with the YOQ total was comparable.

Two-Factor Fixed Exploratory Factor Analysis-Children Time 1

As compared to the previous factor analyses conducted in Winarick, 2018, one of our main new focuses was on longitudinal comparison because we were interested in examining measurement invariance over time. Therefore, we were more rigorous in ensuring that the YOQ used in the analysis fell within 0-4 weeks of an individual's first recorded appointment to be included in Time 1 and 8-12 weeks into treatment to be included in Time 2. As a result, some of the YOQ-SR and YOQ-PRs that were used in the original analyses were discarded for the analyses reported here because they failed to meet the inclusion criteria for each time point.

The common item structure obtained in the initial analyses (Winarick, 2018) was clearly replicated in the new Time 1 exploratory factor analyses detailed here with two minor exceptions. First, two items that were included in the common INT scale in the original analyses ("I don't have friends or I don't keep friends very long," and "I have nightmares, trouble getting to sleep, or waking up too early") no longer met threshold to be included in the YOQ-PR Internalizing scale at Time 1, although they were still included on the INT scale for children at Time 1. Second, one item that was only included on the Parent EXT scale in the original analyses ("I act without thinking and

don't worry about what will happen") now met threshold criteria to be included in the EXT scale for both parents and children at Time 1. The overall consistency of the factor analyses across the two datasets supported the continued use of the initial INT and EXT scales (Table 4).

Several well-recognized criteria for the factorability of a correlation matrix were used. First, the Kaiser-Meyer-Olkin measure of sampling adequacy was .79, above the commonly recommended value of .6, and the communalities were all above .3 (see Table 13), further confirming that each item shared some common variance with other items. Given these overall indicators, factor analysis was deemed to be suitable with all 27 items. Again, three items were removed during previous analyses because they did not load significantly onto either factor in both the child and parent analyses. We used principal axis factoring with an oblique (promax) rotation because the items were assumed to share a common overall symptom factor that would lead to correlated factors. Initial unrotated eigenvalues indicated that the first two factors explained 33% and 12% of the variance respectively (Factor 1 eigenvalue=8.89; Factor 2 eigenvalue=3.32) (Table 10), whereas the remaining factors accounted for far less variance. Examination of the scree plot was also consistent with a two-factor solution and the two factors had clear interpretations consistent with our expectations. Thus, the hypothesized two-factor solution was supported (Table 9).

One item was eliminated because it did not contribute to a simple factor structure. This item was: "I complain or question rules, expectations, or responsibilities." The factor loading matrix for this final solution is presented in Table 13. The factor labels of "internalizing symptoms" and "externalizing symptoms" suited the extracted factors and

were retained. Internal consistency for each of the item sets (Cronbach's alpha) were .91 for Internalizing Symptoms (17 items) and .89 for Externalizing Symptoms (9 items).

Two-Factor Fixed Exploratory Factor Analysis-Parents Time 1

We followed the same procedures for conducting the exploratory factor analyses on the YOQ data for parents. The Kaiser-Meyer-Olkin measure of sampling adequacy was .81, again, above the commonly recommended value of .6, and the communalities were all above .3 (Table 15), further confirming that each item shared some common variance with other items. Given these overall indicators, factor analysis was deemed to be suitable with all 27 items. Similar principal axis factoring and promax rotation methods were used for the YOQ-PR data as were used for the exploratory factor analysis of the YOQ-SR data. Initial unrotated eigenvalues indicated that the first two factors explained 40% and 9% of the variance respectively (Factor 1 eigenvalue=10.72; Factor 2 eigenvalue=2.43) (Table 10).

Five items were eliminated because they did not contribute to a simple two-factor structure. These items were: "My child doesn't participate in activities that used to be fun," "My child's emotions are strong and change quickly," "My child doesn't have friends or he/she does not keep friends very long," "My child has nightmares, trouble getting to sleep, oversleeping, or waking up too early," and "My child feels irritated." The factor loading matrix for this final solution is presented in Table 15. Similar to the child EFA, the factor labels of "internalizing symptoms" and "externalizing symptoms" suited the extracted factors and were retained (Table 4). Internal consistency for each of the item sets (Cronbach's alpha) were .84 for INT (12 items) and .90 for EXT (10 items).

Item level analysis Parents vs. Children at Time 1

The exploratory factor analyses comparing children vs. parents at Time 1 highlighted several items that did not load similarly onto either the internalizing or the externalizing factor. Two items loaded onto child INT and parent EXT at Time 1, indicating that the two groups conceptualize these items as belonging to different sets of symptomatology. Those items include "I have a hard time finishing my assignments or I do them carelessly." and "I have a hard time concentrating, thinking clearly, or sticking to tasks." Five items loaded onto child INT at Time 1 and did not fall onto either scale for parents at Time 1. Those items include "I don't participate in activities that used to be fun," "My emotions are strong and change quickly," "I don't have friends or I don't keep friends very long," "I have nightmares, trouble getting to sleep, oversleeping, or waking up too early," and "I feel irritated." Lastly, one item fell onto parent EXT and did not fall onto either scale for children at Time 1 ("I complain about or question rules, expectations, or responsibilities" (Table 4).

Confirmatory Factor Analyses (CFA) of the Two-Factor Solution on the Common Items in Parents and Children Time 1

Configural Invariance Analyses. We first considered the least constrained two-group model to establish basic structural equivalence between the parents and children at Time 1 (configural invariance) for analyses. To do this, we specified the same two-factor model in each group, but allowed to the estimation of the factor loadings separately for both the children and parents. The fit of the two-group unconstrained configural model was marginally acceptable with an RMSEA=.10 and a CFI=0.83. Although neither of these fit indices are considered good, they are in minimally acceptable ranges.

Metric Invariance Analyses. Therefore, we proceeded to determine if metric invariance existed within this item set. Metric invariance builds upon the assumptions of configural invariance; however, it also constrains factor loadings for items across administrations. To do this, we compared the fit of the metric model to the fit of the configural model using a chi-square difference test. The difference between these two models was significant ($\chi 2=69.07$, df=38, $p<.01$) indicating that the more constrained model fit significantly worse than the less constrained model.

The overall fit of the two-group partially constrained model was again marginal, with an RMSEA=.10 and a CFI=.81. Taken together, the results of our configural and metric invariance tests reflected that, if unconstrained, the items did not load similarly on the INT and EXT factors when comparing parents versus children on the YOQ at Time 1. Additionally, the results of the test for metric invariance indicate that when differences between these two groups arise, it is not necessarily due to differences in the underlying construct. Metric invariance was not established within this item set.

Item Response Theory Analyses (IRT)-Children and Parents at Time 1

Exploratory factor analysis results from comparisons of factor structure between parents and children at Time 1 and Time 2 found that there were clearly structural differences in how parents and children conceptualize items on the YOQ, for instance, they do not even put some of the same items on the same factors (Table 4). However, the purpose of conducting additional IRT analyses was to determine if that lack of measurement invariance was consequential in impacting how individuals respond by examining differential item functioning within the YOQ-SR and YOQ-PR responses. Therefore, in light of the finding that there was basic disagreement about what items

reflected internalizing and externalizing symptomatology, we included all 30 items in our analyses that examined differential item functioning within the YOQ-SR and YOQ-PR.

To conduct graded response analyses through IRT, the data must first meet two basic assumptions: 1) unidimensionality and 2) local dependence. To test the unidimensionality assumption, we conducted an EFA on the full YOQ-SR and YOQ-PR response set at Time 1. Unlike the previous EFA analyses, we allowed all items to load freely instead of restricting the loadings using the two-factor fixed model previously implemented to capture the INT and EXT subscales.

EFA results from the YOQ-PR at Time 1 estimated a Kaiser-Meyer-Olkin measure of sampling adequacy of .76, above the commonly recommended value of .6, and found the communalities were all above .3 (see Table 17). Initial unrotated eigenvalues indicated that the factors explained 40% and 9% of the variance respectively (Factor 1 eigenvalue=10.72; Factor 2 eigenvalue=2.43) (Table 12). EFA results from the YOQ-SR at Time 1 estimated a Kaiser-Meyer-Olkin measure of sampling adequacy of .79 and found all communalities were above .3. Initial unrotated eigenvalues indicated the factors explained 31% and 12% of the variance respectively (Factor 1 eigenvalue=9.38; Factor 2 eigenvalue=3.46) (Table 12). Initial comparisons of the unrestricted YOQ-PR and YOQ-SR EFAs at Time 1 indicated that parents perceive a more unidimensional structure than adolescents.

The local dependence assumption tests the existence of one latent trait and that the participant's responses to YOQ items are not statistically related to each other when that latent trait is held constant. To test this second IRT assumption, we examined the matrix of chi-square values that tested pairwise local independence. We found several

item pairs that strictly violated the local independence assumption of IRT, and the model fit indices were not optimal as a result. Because our focus was on the content of the items and the differences between parents and children, we did not exclude these items from the models but instead note and discuss the substantive implications of these item pairs failure to meet the local independence assumption.

To run the IRT analysis, we had to make several modifications to the dataset to ensure that all items had the same number of responses (Table 2). When examining the child vs. parent comparison at Time 1, the pattern of these slight modifications indicates that children more frequently used higher level endorsements on item-responses than parents. That is, most of the changes to the dataset had to be made by reducing the child endorsement from a 4 to a 3 with the exception of one item (YOQ16).

We then conducted a step-by-step test for differential item functioning (DIF) by following the procedures outlined in the *Methods* section of this paper. There were no items that showed significant DIF (Table 25). Group and item parameter estimates for the parents vs. children comparison at Time 1 are outlined in Table 18 and Tables 23-24, respectively.

Children vs. Parents at Time 2

Initial Tests of Difference

T-Tests. We then compared the Time 2 observed scores of parents and children on the YOQ Total, INT subscale, and EXT subscale. Once again, we found similarity between parent and child reports on the YOQ Total score at Time 2. That is, we did not find a significant difference in the scores for children (M=18.98, SD=14.78) and parents (M=22.22, SD=15.83); $t(135)=1.24$ $p=.23$ on the observed scores at Time 2. These

results suggest that children and parents of children do not differ on their observed reports of overall clinical distress several weeks into psychotherapy treatment.

Furthermore, when comparing scores on the INT Subscale, we also did not find a significant difference for children (M=5.79, SD=5.42) and parents (M=5.75, SD=4.83); $t(135)=-.05$, $p=.96$. However, we did find a significant difference between children (M=3.36, SD=3.46) and parents on the EXT subscale (M=4.95, SD=4.89); $t(135)=2.22$, $p=.03$. These results suggest that although parents and children do not seem to differ on overall levels of distress as indicated by the total score on the YOQ, they do show significant differences in their scores on externalizing symptomatology several weeks into treatment.

Correlations. Correlations between both children and parent's scores on the YOQ Total and INT subscale, YOQ Total and EXT subscale, and INT and EXT subscales were all significant (Table 6). Once again, we found that the INT subscale for the children was more highly correlated with the YOQ Total than the EXT subscale. For the parents, we found that the correlations of the EXT and INT subscales with the YOQ total was comparable, as it was at Time 1.

Two-Factor Fixed Exploratory Factor Analysis-Children Time 2

We followed the same procedures outlined at Time 1 to compare children vs. parents at Time 2, including the removal of the three items that did not load significantly onto either factor in the previous child and parent analyses (Winarick, 2018). For the children at Time 2, the Kaiser-Meyer-Olkin measure of sampling adequacy was .77, and the communalities were all above .3 (Table 14). Initial unrotated eigenvalues indicated that the first two factors explained 35% and 12% of the variance respectively (Factor 1

eigenvalue=9.43; Factor 2 eigenvalue=3.28) (Table 10). Examination of the scree plot was also consistent with a two-factor solution.

Three items were eliminated because they did not contribute to a simple factor structure. These items were: "I think that others are trying to hurt me even when they're not," "My stomach hurts or I feel sick more than others my same age," and "I destroy property on purpose." The factor loading matrix for this final solution is presented in Table 14. Internal consistency for each of the item sets (Cronbach's alpha) were .93 for Internalizing Symptoms (15 items) and .85 for Externalizing Symptoms (9 items).

Two-Factor Fixed Exploratory Factor Analysis-Parents Time 2

The Kaiser-Meyer-Olkin measure of sampling adequacy was .79, again, above the commonly recommended value of .6, and the communalities were all above .3 (see Table 16). Given these overall indicators, factor analysis was deemed to be suitable with all 27 items. Initial unrotated eigenvalues indicated that the first two factors explained 38% and 10% of the variance respectively (Factor 1 eigenvalue=10.13; Factor 2 eigenvalue=2.64).

Three items were eliminated because they did not contribute to a simple two-factor structure. These items were: "My child feels irritated," "My child thinks about suicide or feels that he/she would be better off dead," and "My child gets angry enough to threaten others." The factor loading matrix for this final solution is presented in Table 16. Internal consistency for each of the item sets (Cronbach's alpha) were .90 for INT (13 items) and .89 for EXT (11 items).

Item level analysis Parents vs. Children at Time 2

Item level analysis of the two-factor fixed EFA at Time 2 indicated several notable findings. Namely, there were two items that loaded significantly onto the parent

INT scale at Time 2, but did not load significantly onto either factor for the children at Time 2. These items were, "I think that others are trying to hurt me even when they're not" and "My stomach hurts or I feel sick more than others my same age." There was also one item that loaded significantly onto the parent EXT scale at Time 2, but did not load significantly onto either factor for the children at Time 2. This item was, "My child destroys property on purpose." Lastly, we found that there were two items that switched factors from the parent EXT scale to the child INT scale at Time 2. These items were, "I have a hard time finishing my assignments or I do them carelessly" and "I have a hard time concentrating, thinking clearly, or sticking to tasks." Interestingly, these were the same two items that switched factors at Time 1 as well when evaluating INT and EXT factor structure for parents vs. children (Table 4).

Confirmatory Factor Analyses (CFA) of the Two-Factor Solution on the Common Items in Parents and Children Time 2.

Configural Invariance Analyses. Similar to the steps followed for the CFA in Parents and Children at Time 1, we first considered the least constrained two-group model to establish basic structural equivalence between the parents and children at Time 2 (configural invariance) for analyses. To do this, we specified the same two-factor model in each group, but allowed to the estimation of the factor loadings separately for both the children and parents. Configural invariance analyses revealed that the fit of the two-group unconstrained model exceeded common acceptability standards (Byrne, 2012), with an RMSEA=.13 and a CFI=0.69. Due to lack of measurement invariance in the least constrained model, we did not proceed with the additional test of metric invariance.

Item Response Theory Analyses (IRT)-Children and Parents at Time 2

EFA results using the full 30-item YOQ-PR at Time 2 estimated a Kaiser-Meyer-Olkin measure of sampling adequacy of .76, above the commonly recommended value of .6, and found the communalities were all above .3 (Table 17). Initial unrotated eigenvalues indicated that the factors explained 36% and 9% of the variance respectively (Factor 1 eigenvalue=10.69; Factor 2 eigenvalue=2.71) (Table 12). EFA results using the full 30-item YOQ-SR at Time 2 estimated a Kaiser-Meyer-Olkin measure of sampling adequacy of .72 and found all communalities were above .3 (Table 17). Initial unrotated eigenvalues indicated the factors explained 33% and 11% of the variance respectively (Factor 1 eigenvalue=9.38; Factor 2 eigenvalue=3.46) (Table 12). Similar to the Time 1 comparisons, the unrestricted YOQ-PR and YOQ-SR EFAs at Time 2 indicate that parents perceive a more unidimensional structure than do adolescents.

As with the child vs. parent comparisons at Time 1, we found several item pairs that strictly violated the local independence assumption of IRT and as a result the model fit indices were not optimal. For the purposes of the child vs. parent comparisons at Time 2, we had to change several mismatched responses in the dataset in order to run the IRT analyses (Table 2). The pattern of the modifications for child vs. parent comparisons at Time 2 was mixed, as changes had to be made to both the child and parent responses to proceed with the analyses.

We then conducted a step-by-step test for differential item functioning (DIF) by following the procedures outlined in the *Methods* section of this paper. We found significant DIF in the following items, “I have physical fights (hitting, kicking, biting, or scratching) with my family or others my age” (p=0.03), “I have nightmares, trouble

getting to sleep, oversleeping, or waking up too early" (p=0.01), and "I break rules, laws, or don't meet others' expectations on purpose" (p=0.05) (Table 28). These findings indicate that parents and children respond to these items differently and may impact how their total scores are perceived in comparison to one another. All three of these items fall onto the same INT and EXT subscales for both children and parents at Time 2, therefore indicating that they view them as in line with similar symptomatology.

Results of the DIF analysis on the item, "I have physical fights (hitting, kicking, biting, or scratching) with my family or others my age," found that children view this as a "harder" item than parents. This finding suggests that children need to have significantly more psychological distress than their parents perceive to endorse a 1 ("Rarely") instead of a 0 ("Never or almost never"), b_1 (children)=1.86, b_1(parents)=.68, and almost five times more distress compared to their parent's perception to endorse a 2 ("Sometimes") instead of a 1 ("Rarely") on the YOQ response scale, b_2 (children)=5.22, b_2(parents)=1.15.

Similarly, children also view the item, "I break rules, laws, or don't meet others' expectations on purpose," as a "harder" item than the parents. This finding suggests that children need to have significantly more psychological distress than their parents perceive to endorse a 1 ("Rarely") instead of a 0 ("Never or almost never"), b_1 (children)=.88, b_1(parents)=.14, and almost double the amount of distress compared to their parent's perception to endorse a 2 ("Sometimes") instead of a 1 ("Rarely") on the YOQ response scale, b_2 (children)=2.11, b_2(parents)=1.01.

Additionally, results of the DIF analysis on the item, "I have nightmares, trouble getting to sleep, oversleeping, or waking up too early," found that children view this as an

"easier" item than parents. This finding suggests that children need to have significantly less psychological distress than their parents perceive to endorse a 1 ("Rarely") instead of a 0 ("Never or almost never"), b_1 (children)=-.75, b_1(parents)=.18; a 2 ("Sometimes") instead of a 1 ("Rarely"), b_2 (children)=.28, b_2(parents)=1.03; a 3 ("Frequently") instead of a 2 ("Sometimes"), b_3 (children)=.1.14, b_3(parents)=1.30; and a 4 ("Almost Always or Always") instead of a 3 ("Frequently") on the YOQ response scale, b_4(children)=2.02, b_4(parents)=3.22. Item and group parameter estimates for the parent vs. child comparison at Time 2 are outlined in Tables 26-27. Graphs of items showing DIF are found in Figure 1.

Children Time 1 vs. Children Time 2

Initial Tests of Difference

T-Tests. We then compared the Time 1 vs. Time 2 observed scores of children on the YOQ Total, INT subscale, and EXT subscale. We found that children reported significantly different overall levels of distress at Time 1 (M=29.85, SD=16.77) compared to Time 2 (M=18.98, SD=14.78); $t(151)=4.24$ $p=<.00$. When we examined specific subscales, we also found that children's observed ratings of INT symptomatology differ significantly at Time 1 (M=8.63, SD=6.25) compared to Time 2 (M=5.79; SD=5.42); $t(151)=3.0$ $p=<.00$. Similarly, children's ratings of EXT symptomatology significantly decrease from Time 1 (M=6.59, SD=4.71) to Time 2 (M=3.36, SD=3.46) $t(151)=4.81$ $p=<.00$.

Correlations. We found higher correlations between INT and YOQ Total than EXT and YOQ Total for children at both Time 1 and Time 2 (Table 7).

Item Level Analysis-Two-Factor Fixed EFA-Children Time 1 vs. Time 2

A comparison of the YOQ-SR two-factor structure over time shows that many of the items remained from Time 1 to Time 2 (Table 4). That is, most items that fell on the INT/EXT scales at Time 1 also remained on the INT/EXT scales at Time 2. There were four exceptions to this finding. Two items, "I think that others are trying to hurt me even when they are not" and "My stomach hurts or I feel sick more than others my same age" fell on the INT scale at Time 1 and did not load significantly onto either factor at Time 2. One item, "I complain about or question rules, expectations, or responsibilities," did not load significantly onto either factor at Time 1, and fell onto the EXT scale at Time 2. Lastly, one item, "I destroy property on purpose," fell onto the EXT scale at Time 1 and did not load significantly onto either scale at Time 2.

Confirmatory Factor Analyses (CFA) of the Two-Factor Solution on the Common Items in Children Time 1 vs. Children Time 2

Configural Invariance Analyses. Similar to the steps followed for the previous CFA analyses, we first considered the least constrained two-group model to establish basic structural equivalence between the children at Time 1 and Time 2 (configural invariance) for analyses. To do this, we specified the same two-factor model in each group, but allowed to the estimation of the factor loadings separately for both the Time 1 group and the Time 2 group. Configural invariance analyses revealed that the fit of the two-group unconstrained model exceeded common acceptability standards (Byrne, 2012), with an RMSEA=.12 and a CFI=0.75. Due to lack of measurement invariance in the least constrained model, we were unable to proceed with additional tests of metric invariance.

Item Response Theory Analyses (IRT)-Children at Time 1 vs. Children at Time 2

Comparison using the full 30-item YOQ-SR at Time 1 and Time 2 found that children remain consistent in EFA factor variance results over time (Table 12). Examination of the matrix of chi-square values that tested pairwise local independence found several item pairs that were considered to have larger associations beyond the single latent trait than would be considered "good fit."

We then conducted a step-by-step test for differential item functioning (DIF) by following the procedures outlined in the *Methods* section of this paper. We found significant DIF in the following item, "I have physical fights (hitting, kicking, biting, or scratching) with my family or others my age" (p=0.01) (Table 31). This finding suggests that children need to have significantly more psychological distress at Time 2 than Time 1 to endorse a 1 ("Rarely") instead of a 0 ("Never or Almost Never"), b_1 (Time 1)=.27, b_1(Time 2)=1.72, and double the distress at Time 2 than Time 1 to endorse a 2 ("Sometimes") instead of a 1 ("Rarely") on the YOQ response scale, b_2 (Time 1)=2.44, b_2(Time 2)=5.71. Item and group parameter estimates for children at Time 1 vs. Time 2 are outlined in Tables 29-30. Graphs of the item showing DIF are found in Figure 2.

Parents Time 1 vs. Parents Time 2

Initial Tests of Difference

T-Tests. We then compared the Time 1 vs. Time 2 observed scores of parents on the YOQ Total, INT subscale, and EXT subscale. Similarly to the child results, we found that parents reported significantly different overall levels of distress at Time 1 (M=29.43, SD=18.32) compared to Time 2 (M=22.22, SD=15.83); $t(131)=2.41$ $p=.02$. When we examined specific subscales, we also found that parent's observed ratings of INT

symptomatology differ significantly at Time 1 (M=7.60, SD=5.33) compared to Time 2 (M=5.75; SD=4.83); $t(131)=2.09$ $p=.04$. However, parent's observed ratings of EXT symptomatology did not significantly decrease from Time 1 (M=6.61, SD=5.55) to Time 2 (M=4.95, SD=4.89) $t(131)=1.82$ $p=.07$.

Correlations. We found comparable correlations between INT and EXT subscales with the YOQ Total for parents at both Time 1 and Time 2 (Table 8).

Item Level Analysis-Two-Factor Fixed EFA-Parents Time 1 vs. Time 2

A comparison of the YOQ-PR two-factor structure over time shows that many of the items remained the same from Time 1 to Time 2 (Table 4). That is, most items that fell on the INT/EXT scales at Time 1 also fell on the INT/EXT scales at Time 2. There were six exceptions to this finding. Four items, "My child doesn't participate in activities that used to be fun," "My child's emotions are strong and change quickly," "My child doesn't have friends or doesn't keep friends for very long," and "My child has nightmares, trouble getting to sleep, oversleeping, or waking up too early," that did not load significantly onto either factor at Time 1 and fell on the INT scale at Time 2. One item, "My child thinks about suicide or feels he/she would be better off dead," loaded onto the INT scale at Time 1 and did not load significantly onto either scale at Time 2. Lastly, one item, "My child gets angry enough to threaten others," fell on the EXT scale at Time 1 and did not load significantly onto either factor at Time 2.

Confirmatory Factor Analyses (CFA) of the Two-Factor Solution on the Common Items in Parents Time 1 vs. Parents Time 2

Configural Invariance Analyses. Similar to the steps followed for the previous CFA analyses, we first considered the least constrained two-group model to establish

basic structural equivalence between the parents at Time 1 and Time 2 (configural invariance) for analyses. To do this, we specified the same two-factor model in each group, but allowed to the estimation of the factor loadings separately for both the Time 1 group and the Time 2 group. Configural invariance analyses revealed that the fit of the two-group unconstrained model exceeded common acceptability standards (Byrne, 2012), with an RMSEA=.12 and a CFI=0.77. Due to lack of measurement invariance in the least constrained model, we were unable to proceed with additional tests of metric invariance.

Item Response Theory Analyses (IRT)-Parents at Time 1 vs. Parents at Time 2

Comparison using the full 30-item YOQ-PR at Time 1 and Time 2 found that parents remain consistent in EFA factor variance results over time (Table 12). Examination of the matrix of chi-square values that tested pairwise local independence found several item pairs that were considered to have larger associations beyond the single latent trait than would be considered "good fit."

We then conducted a step-by-step test for differential item functioning (DIF) by following the procedures outlined in the *Methods* section of this paper. We found significant DIF in the following items, "My child has a hard time sitting still (or he/she has too much energy)" (p=0.02) and "My child has nightmares, trouble getting to sleep, oversleeping, or waking up too early" (p=0.02) (Table 34).

Results of the DIF analysis on the item, "My child has a hard time sitting still (or he/she has too much energy)" found that parents view this as an "easier" item at Time 2 than Time 1. This finding indicates that parents need to perceive less psychological distress in their children at Time 2 to endorse higher symptoms ratings compared to Time

1. Conversely, for the item "My child has nightmares, trouble getting to sleep, oversleeping, or waking up too early," DIF analyses revealed that parents need to perceive more distress at Time 1 compared to Time 2 to endorse a 1 ("Rarely") instead of a 0 ("Never or Almost Never"), b_1 (Time 1)=-.64, b_1(Time 2)=-.24. However, this pattern changes as higher ratings are taken into account. It then requires parents to perceive significantly more distress in their children at Time 2 compared to Time 1 to endorse higher ratings on the YOQ response scale (Table 34).

These findings indicate that parents respond differently to these items over time and may impact how their total scores are perceived in comparison to one another. Item and group parameter estimates for the parent Time 1 vs. Time 2 comparison are outlined in Tables 32-33. Graphs of items showing DIF are found in Figure 3.

Discussion

Our results paint a complicated picture of how parents and children respond to items on the YOQ, both between participant groups at each time point (parent vs. child) and within participant groups across time (e.g., child T1 vs. child T2). There are clearly differences between how parents' and children's responses to YOQ items relate to the underlying latent variables of internalizing and externalizing symptomatology.

There is a breadth of research that exists on multi-informant discrepancies between parents and children. One of the potential reasons for these discrepancies is the differences in perception of internalizing and externalizing symptoms (De Los Reyes et al., 2015); specifically, that children are inclined to view symptoms as the result of internalizing disorders, whereas parents view symptoms as reflective of externalizing disorders. Our hypotheses followed from this observation.

Parent and Child Perceptions at the Start of Therapy

We hypothesized that our exploratory factor analyses conducted on child and parent comparisons at Time 1 would find that different items loaded onto different factors and that items that switch factors will do so in the direction of Child INT to Parent EXT (***Hypothesis 1a***). Results from the current Time 1 exploratory factor analyses largely confirmed this hypothesis. What, then, are the implications of these findings? Clinically, our EFA results may help identify specific symptomatology that parents and children disagree about when it comes to assessing presenting psychological problems at baseline.

For example, two items loaded onto the Child INT scale and Parent EXT scale at Time 1, indicating that the two groups conceptualize these items as belonging to different sets of symptomatology when starting treatment. Those items include "I have a hard time

finishing my assignments or I do them carelessly." and "I have a hard time concentrating, thinking clearly, or sticking to tasks." These items are markers of inattentiveness and are in line with a symptom profile of attention-deficit/ hyperactivity disorder (ADHD; DSM-5; American Psychiatric Association, 2013); however, they are *also* in line with the inattentiveness exhibited in children and adolescents with anxiety, depression, and histories of trauma (2013).

Although ADHD has been intensively studied, much less is known about the way potentially traumatic experiences influence attentional deficits in children and adolescents. Results from one study with over 76,000 children aged 4 to 17 years found that children with diagnoses of ADHD also had a much higher prevalence of adverse childhood experiences (ACEs) (Brown et al., 2017). These faulty, or incomplete, interpretations of children's behavior can impact the way parents speak to clinicians about their child's symptoms and treatment progress and can certainly influence how they respond to measures of child mental health such as the YOQ.

Inattentiveness is also a common symptom of anxiety disorders, such as generalized anxiety disorder and social phobia (American Psychiatric Association, 2013). Anxiety disorders, trauma, and ADHD have similar symptomatology but vary in terms of treatment approaches and potential psychopharmacological recommendations. Researchers estimate that the comorbidity between an anxiety disorder and ADHD is roughly 30% (Larson et al., 2011). Therefore, the misdiagnosis of one for the other based on a single-informant assessment has important clinical implications. In one study conducted with 230 child psychologists, researchers found that only 15% of them reported using multiple methods of assessment before making an ADHD diagnosis

(Handler, 2005). Our results show that if a clinician makes this assessment using just a parent-report, then that clinician is in danger of misdiagnosing or not assessing for other potential reasons for the child's problems, such as an underlying anxiety disorder.

However, it is also important to note that the wording of the YOQ asks parents to rate symptoms that they observe in their adolescents and asks adolescents to rate symptoms that they experience. This differentiation between observed behaviors and internal experiences can also attribute to the differences in perceived inattentiveness because the only time that parents may notice their child being inattentive is in the context of other externalizing symptoms. Therefore, it is also possible that this discrepancy in factor loadings may occur because parents lack the ability to identify inattentiveness when it is more in line with an internalizing disorder and therefore only endorse it as a symptom of other more observable behaviors.

Also, the differences in factor structure between parents and children speak to a potentially higher degree of comorbidity between symptom type and lend itself to a more dimensional, rather than categorical, diagnostic approach. The DSM-5 has been commonly criticized for its categorical approach to diagnostic criteria and our results demonstrate how the overlap in symptomatology warrant further discussion about the use of dimensional classifications as they may be more valid systems of diagnosis. Therefore, clinicians must be cautious when making treatment recommendations and diagnostic interpretations simply by comparing observed scores on these measures.

When examining differences between parent and child reporting of internalizing and externalizing symptoms at Time 1, we also found five items that loaded onto the child INT and did not fall onto either scale for parents at Time 1. Those items include "I

don't participate in activities that used to be fun," "My emotions are strong and change quickly," "I don't have friends or I don't keep friends very long," "I have nightmares, trouble getting to sleep, oversleeping, or waking up too early," and "I feel irritated." Interestingly, most of these items focus on behavioral and emotional symptoms that commonly underly internalizing disorders such as anxiety or depression (American Psychiatric Association, 2013). The fact that they did not fall on either scale for Parents at Time 1 may indicate that parents are less aware of these behaviors as clinical markers of internalizing disorders such as anxiety, depression, and/or trauma. These findings were in line with the confirmation of our hypothesis that the two-factor structure did fit better for children than parents at Time 1 as evidenced by a lower eigenvalue ratio of the first and second factors (***Hypothesis 1b***). Understanding this discrepancy in the initial perception of symptoms may aide both parents and clinicians in using these behaviors as helpful indicators for internalizing disorders.

Although we did not find strong evidence for measurement invariance between the parents and children at Time 1, the IRT analysis also did not suggest that there was much differential item functioning. In fact, we failed to find DIF on any items in the parent and child comparison at Time 1 **(*Hypothesis 1c*)**. This finding implies that although children and parents view the items on the YOQ as belonging to different sets of symptoms at baseline, they do not differ on the amount of latent psychological distress required to endorse specific responses. Not only did we have a lack of difference between parents and children at Time 1 on latent distress, but we also found no difference between parents and children on the YOQ total observed score. Clinically, this finding

helps bolster evidence for the continued use of the observed YOQ-SR/PR total score as similarly indicative of overall psychological distress when starting treatment.

Parent and Children Perceptions Later in Therapy

We correctly hypothesized that we would find more agreement between parents and children on INT and EXT item loadings at Time 2 compared to Time 1 ***(Hypothesis 2a)***. For example, four of the five items that fell onto Child INT at Time 1 but did not fall onto either factor for parents at Time 1, now fell onto the Parent INT scale at Time 2. This finding indicates a greater agreement between parents and children of internalizing symptomatology several months into treatment compared to the start of treatment.

One of the most vital components of the evidence-based treatment of children and adolescents is caregiver involvement and psychoeducation (Sauter et al., 2009; Barmish & Kendall, 2005; Kendall, 2000). It is through these collateral sessions that the treating clinician can provide information about symptoms and treatment, collaboratively develop treatment goals, and answer questions (Sauter et al., 2009). Given that adolescents may spend the majority of their time in school or with peers, these sessions also serve to help develop more effective communication between the adolescent and their caregiver (Stallard, 2009; Kingery et al., 2006). Hence, it is possible that this increased agreement between parents and children on INT symptoms at Time 2 may be a result of these factors that typically occur within the first few sessions of treatment.

When we examined the within-group Time 1 and Time 2 comparisons, we found that many of the items on the two-factor structure remained invariant both across and within the respondent group over time, that is, Parent Time 1 vs. Parent Time 2, Child Time 1 vs. Child Time 2, Parent vs. Child Time at 1 and Parent vs. Child at Time 2.

There were 14 items (from the 27-item set) that remained consistent across YOQ-PR/SR at Time 1 and Time 2, 7 items on the INT scale and 7 items on the EXT scale. This finding suggests that clinicians and researchers may want to pay particular attention to these items as illustrative of stable perceptions of internalizing and externalizing symptoms in both parents and children over time.

Furthermore, the EFAs at Time 2 (8-12 weeks into treatment) found that there were two items that switched factors between YOQ-SR and YOQ-PR responses, and did so in the same direction: from Child INT to Parent EXT. These items were, "I have a hard time finishing my assignments or I do them carelessly" and "I have a hard time concentrating, thinking clearly, or sticking to tasks." These were the same two items that switched factors at Time 1 when evaluating INT and EXT factor structure for parents vs. children. The consistency in the longitudinal factor structure may indicate a stronger perceived difference of symptoms of inattentiveness. Analyses failed to replicate similar findings in an item measuring hyperactivity, "I act without thinking and don't worry about what will happen." Again, this discrepancy becomes apparent in the clinical literature when making a differential diagnosis between ADHD and an anxiety disorder.

We also confirmed our hypothesis that the two-factor structure would fit better for children compared to parents at Time 2 as evidenced by the smaller eigenvalue ratio ***(Hypothesis 2b)***. These results indicate that although children and parents show more agreement on their conceptualizations of INT and EXT symptoms at Time 2, children still differentiate their symptomatology in their responses to the YOQ more so than their parents. This differentiation provides important clinical information because it shows that children and adolescents may be more able to effectively distinguish their symptoms,

whereas parents are more likely to present to a clinic with a general sense that something is wrong, but lack the acuity to aptly identify exactly what that "something" is.

We found additional discrepancies in the understanding of INT and EXT symptoms at Time 2 when we examined the results of the DIF analyses. Unlike our results at Time 1, there were three items at Time 2 that exhibited significant DIF and disconfirmed our hypothesis that there will be less differential item functioning between YOQ-SR and YOQ-PR analyses than at Time 1 ***(Hypothesis 2c).*** Those items were, "I have physical fights (hitting, kicking, biting, or scratching) with my family or others my age," "I break rules, laws, or don't meet others' expectations on purpose," and "I have nightmares, trouble getting to sleep, oversleeping, or waking up too early."

The first two items that showed DIF fell onto the EXT factor for children and parents at both Time 1 and Time 2 and were rated as more "difficult" items for the children to endorse compared to their parents. That is, the parents required less perceived underlying EXT symptoms in their children at Time 2 to endorse higher responses for each of these items. The last of the three items fell onto the INT factor for both respondents across each time point and was rated as "easier" for the children to endorse compared to their parents. This difference shows that the children required less underlying INT symptoms at Time 2 to endorse higher responses on this item.

Taken together, these findings correspond with previous literature, which states that informant discrepancy may occur as a function of the types of symptoms (De Los Reyes et al., 2015; Achenbach et al. 1987). Parents tend to be more aware of observable, externalizing behaviors such as those stated in the items that demonstrated DIF (i.e., physical fighting and rule-breaking). There is also literature that suggests informant

discrepancies in externalizing behavior may partly be due to a child or adolescent's lack of self-awareness (Owens & Hoza, 2003). Once again, the results from these latent trait analyses corresponded with significant differences found in observed scores on the YOQ at Time 2. T-tests found that children had significantly lower scores on the EXT scale compared to parents. The IRT results suggest that differential item functioning in two of the seven items on the EXT scale may cause this discrepancy because children require significantly more latent distress than parents perceive to endorse the same rating on the YOQ response scale.

Changes Within Informants Over Time

Another important clinical finding from this study is that we largely confirmed our hypotheses when examining the within-group comparison for children across time points. We hypothesized that the factor structure of INT and EXT subscales would remain invariant over time ***(Hypothesis 3a).*** Results predominantly confirmed this assumption with a few minor exceptions. For example, EFA results showed that two items loaded on to the INT scale at Time 1 and no longer met the threshold for inclusion on either scale at Time 2. These items were, "I think that others are trying to hurt me even when they are not" and "My stomach hurts or I feel sick more than others my same age." Furthermore, we found that one item, "I complain about or question rules, expectations, or responsibilities," did not load significantly onto either factor at Time 1 and loaded onto the Externalizing scale at Time 2. Lastly, one item, "I destroy property on purpose," fell onto the Externalizing scale at Time 1 and did not load significantly onto either scale at Time 2.

Overall, we found that children's views of their distress shifted slighted towards being in line with their parent's views of EXT symptomatology at Time 2, although they maintained roughly the same eigenvalue ratio as they had at Time 1 ***(Hypothesis 3b***). One potential reason for the gradual shift in the children's rating of these items is an increase in self-awareness. A common treatment factor amongst evidence-based treatments for children and adolescents is an emphasis on self-monitoring of thoughts, feelings, physiologic sensations, and behaviors (Chorpita & Weisz, 2009). Unfortunately, we did not have intervention data from individual therapy sessions, therefore future research warrants investigation of the link between increased self-perception/endorsement of externalizing symptoms and specific treatment interventions.

Thankfully, observed scores on total distress, INT, EXT subscales significantly decreased from Time 1 to Time 2. We were able to infer that this change was, in fact, linked to a change in latent levels of distress given the substantial lack of DIF in child comparisons across time ***(Hypothesis 3c).*** There was one exception, "I have physical fights (hitting, kicking, biting, or scratching) with my family or others my age" that showed significant DIF. Children need to have significantly more psychological distress at Time 2 than Time 1 to endorse higher responses on this item.

Overall, the self-reported decrease in symptoms may provide evidence for the effectiveness of psychotherapeutic treatment. We performed this study at a community mental health center where clinicians are trained to treat children and adolescents using evidence-based modalities, such as cognitive-behavioral therapy (CBT). There is a vast literature supporting the use of CBT for both internalizing (Oud et al., 2019; Stikkelbroek et al., 2013; Sauter et al., 2009) and externalizing (Battagilese et al., 2015; Lochman et

al., 2011; Weisz, Hawley, & Doss, 2004) disorders. Results from the comparisons of children's YOQ scores across time provide additional support of this literature, although future research on the impact of specific treatment interventions is necessary.

Finally, in examining the longitudinal differences between parent reports on the YOQ, we found that more items loaded onto the INT scale at Time 2 and there was clearer distinguishing of items as belonging to either INT or EXT ***(Hypothesis 4a).*** Interestingly, it appears as though parents were better able to identify the INT symptoms that they missed at baseline, such as not engaging in previously enjoyed activities, emotional dysregulation and irritability, and sleep difficulties. Clinically, we know that these are common symptoms of childhood depression (American Psychiatric Association, 2013); but children express them in a way that many parents may view as a result of their child being oppositional or "difficult." The results of the longitudinal EFA demonstrated a change in the perception of these symptoms and may speak to the effectiveness of parent psychoeducation on internalizing symptomatology.

Our hypothesis that the two-factor structure would fit better at Time 2 was disconfirmed as evidenced by the similar eigenvalue ratio at Time 1 and Time 2 ***(Hypothesis 4b).*** However, we did find minor evidence that supported our hypothesis that there would be significant differential item functioning between YOQ-PR Time 1 and Time 2 ***(Hypothesis 4c)***. The IRT analysis on the item "My child has a hard time sitting still (or he/she has too much energy)" indicated that parents need to perceive less psychological distress in their children at Time 2 to endorse higher symptom ratings compared to Time 1. Conversely, for the item "My child has nightmares, trouble getting to sleep, oversleeping, or waking up too early," DIF analyses revealed that parents need

to perceive more distress at Time 1 compared to Time 2; however, this pattern changes as we account for higher ratings.

When we examined the overall observed means of the YOQ-PR Total score, INT scale, and EXT scale, we found that parents perceived less overall distress and internalizing symptoms as their child progresses through treatment. Results from our IRT analyses allowed us to interpret these findings as indicative of clinically significant change given the broad lack of DIF on the YOQ-PR. Clinically, the lack of DIF supports the continued use of the YOQ as an informative instrument to assess changes in psychological distress throughout treatment.

Limitations and Future Directions

Our study had several limitations, including the substantial evidence of a lack of local independence between items when testing the assumptions to interpret IRT results. This lack of local independence is perhaps not surprising because the YOQ is intended to measure a broad set of symptoms, some of which will be more similar to others than other sets, such as INT items and EXT items. However, there is also strong evidence that we can treat the YOQ as a unidimensional scale (Wells, Burlingame, & Rose, 2003; Dunn et al., 2005). A unidimensional scale by definition shouldn't have a lot of local dependencies because there is only one factor. Therefore, we chose to focus on the unidimensionality assumption for IRT, which was generally met in our assessments of the YOQ across participants and points of administration.

A second limitation was the relatively small sample size and lack of demographic information regarding participant sex and ethnicity. Additional research including more participants would allow for greater generalizability of our results. Furthermore, there is

ample research on sex differences in rates of mental health disorders in the child and adolescent population. For example, Van Droogenbroeck et al. (2018) examined gender differences in mental health problems amongst adolescents and young adults and found that girls reported significantly higher levels of overall psychological distress and internalizing disorders, such as depression and anxiety, compared to same-aged male peers. Conversely, epidemiologic research on externalizing disorders has found that men exhibit higher prevalence rates than women (Hicks et al., 2007). Additional information on participant sex is warranted given these differences in prevalence and reporting rates of mental health concerns.

Views of mental health symptoms also vary across different ethnicities and cultural backgrounds. Several factors impact cultural identity, including social norms, customs, and religious frameworks. Culture shapes the way in which people view themselves as well as how they view mental health disorders. For example, Haque (2010) found that shame influences the expression of mental health concerns to family members in individuals from Southeast Asian countries. The cultural background of the client also influences the way that clinicians perceive their mental health symptoms (Biswas, Gangadhar, & Keshavan, 2016). Therefore, it is important that future research consider a respondent's cultural background when assessing for evaluation of symptom change throughout the course of psychotherapy.

A third limitation was our lack of information on parental mental health, particularly maternal depression. One study found maternal depression to be the strongest predictor of parent-child agreement on child mental health symptoms (Popp et al., 2017). Additional research shows that maternal anxiety and depression influenced

the way in which mothers rated their child's internalizing symptoms (Kelley et al., 2017). Therefore, future studies examining the role of parental mental health on perceptions of child and adolescent symptomatology would be beneficial in the larger understanding of multi-informant discrepancies.

Lastly, our study did not include information on the number of sessions attended between time points or on the psychotherapeutic interventions employed in each session. Including this information would allow researchers to more accurately draw conclusions about the effectiveness of psychotherapy in helping to reduce the rates of psychological distress. Information on number of sessions and intervention data would also add to the understanding of the specific factors that influence agreement or disagreement on symptomatology over time.

Conclusion

Our research highlights the importance of gaining a comprehensive understanding of the ways in which children and their caregivers perceive symptoms of mental health disorders. Although multi-informant data is often discrepant, it provides valuable clinical information that aides clinicians and researchers in formulating diagnoses, developing treatment plans, and tracking change over time. The results from our study show that children and parents perceive symptoms differently, especially behavioral indicators of internalizing disorders such as anxiety and depression. Therefore, it is vital that the treating clinician understand and acknowledge these discrepancies when developing an assessment-based case conceptualization and collaborative treatment plan.

Tables

Table 1

Description of Participants

	N	*Mean Age of Child*	*Min. Age*	*Max. Age*	*St. Dev. Age*	*Mean Weeks in Tx*	*St. Dev. Weeks in Tx*
Child T1	79	14.58	12	17	1.68	1.00	1.66
Parent T1	70	14.70	12	17	1.61	0.60	1.38
Child T2	74	14.15	11	17	1.88	10.30	1.17
Parent T2	63	13.83	12	17	1.96	10.02	1.23

Table 2

Item Response Modifications for IRT Analyses

Item	*Parent vs. Child T1*	*Parent vs. Child T2*	*Child T1 vs. Child T2*	*Parent T1 vs. Parent T2*
YOQ-1	1 child endorsed a 4, needed to change to a 3	----	1 person endorsed a 4, needed to change to a 3	----
YOQ-3	----	1 parent endorsed a 4, needed to change to a 3	4 people endorsed a 4, needed to change to a 3	----
YOQ-6	----	3 parents endorsed a 3, needed to change to a 2	2 people endorsed a 3, 2 people endorsed a 4, needed to change all to a 2	2 people endorsed a 4, needed to change to a 3
YOQ-8	----	----	1 person endorsed a 4, needed to change to a 3	1 person endorsed a 4, needed to change to a 3
YOQ-10	----	----	1 person endorsed a 3, needed to change to a 2	1 person endorsed a 3, 1 person endorsed a 4, needed to change all to a 2
YOQ-11	----	2 children endorsed a 3, needed to change to a 2	3 people endorsed a 4, needed to change to a 3	4 people endorsed a 3, 1 person endorsed a 4, needed to change all to a 2
YOQ-12	----	3 children endorsed a 4, needed to change to a 3	----	2 people endorsed a 4, needed to change to a 3
YOQ-14	1 child endorsed a 4, needed to change to a 3	----	1 person endorsed a 4, needed to change to a 3	----

YOQ-15	----	----	5 people endorsed a 4, needed to change to a 3	2 people endorsed a 3, needed to change to a 2
YOQ-16	1 parent endorsed a 4, needed to change to a 3	2 parents endorsed a 3, needed to change to a 2	1 person endorsed a 4, needed to change to a 3	1 person endorsed a 4, needed to change to a 3
YOQ-17	3 children endorsed a 4, needed to change to a 3	----	3 people endorsed a 4, needed to change to a 3	----
YOQ-19	----	1 child endorsed a 4, needed to change to a 2	1 person endorsed a 4, needed to change to a 3	3 people endorsed a 3, 1 person endorsed a 4, needed to change all to a 2
YOQ-22	----	5 parents endorsed a 3, needed to change to a 2	3 people endorsed a 3, 2 people endorsed a 4, needed to change all to a 2	1 person endorsed a 4, needed to change to a 3
YOQ-24	----	1 child endorsed a 4, needed to change to a 3	----	3 people endorsed a 4, needed to change to a 3
YOQ-25	----	----	2 people endorsed a 4, needed to change to a 3	2 people endorsed a 4, needed to change to a 3
YOQ-26	----	2 parents endorsed a 3, needed to change to a 2	2 people endorsed a 3, 1 person endorsed a 4, needed to change all to a 2	1 person endorsed a 4, needed to change to a 3
YOQ-28	----	2 children endorsed a 4, needed to change to a 3	----	4 people endorsed a 4, needed to change to a 3
YOQ-30	2 children endorsed a 4, needed to change to a 3	1 child endorsed a 4, needed to change to a 3	----	----

Table 3

YOQ Items by Factor (Winarick, 2018)

YOQ Item	Child INT	Parent INT	Child EXT	Parent EXT	No INT/EXT-Child	No INT/EXT-Parent	Switched factors
1. *I have headaches or feel dizzy.*	✔	✔					
2. *I don't participate in activities that used to be fun.*	✔					✔	
3. *I argue or speak rudely to others.*			✔	✔			
4. *I have a hard time finishing my assignments or I do them carelessly.*	✔			✔			✔
5. *My emotions are strong and change quickly.*	✔			✔			✔
6. *I have physical fights (hitting, kicking, biting, or scratching) with my family or others my age.*			✔	✔			
7. *I worry and can't get thoughts out of my mind.*	✔	✔					
8. *I steal or lie.*			✔	✔			
9. *I have a hard time sitting still (or I have too much energy).*					✔	✔	
10. *I use alcohol or drugs.*					✔	✔	
11. *I am tense and easily startled (jumpy).*	✔	✔					
12. *I am sad or unhappy.*	✔	✔					
13. *I have a hard time trusting friends, family members, or other adults.*	✔	✔					
14. *I think that others are trying to hurt me even when they are not.*	✔	✔					
15. *I have threatened to, or have run away from home.*					✔	✔	
16. *I physically fight with adults.*			✔	✔			
17. *My stomach hurts or I feel sick more than others my age.*	✔	✔					
18. *I don't have friends or I don't keep friends long.*	✔	✔					
19. *I think about suicide or feel I would be better off dead.*	✔	✔					
20. *I have nightmares, trouble getting to sleep, oversleeping, or waking up too early.*	✔	✔					
21. *I complain about or question rules, expectations, or responsibilities.*	✔			✔			✔
22. *I break rules, laws, or don't meet others' expectations on purpose.*			✔	✔			
23. *I feel irritated.*	✔					✔	
24. *I get angry enough to threaten others.*			✔	✔			
25. *I get into trouble when I'm bored.*			✔	✔			
26. *I destroy property on purpose.*			✔	✔			
27. *I have a hard time concentrating, thinking clearly, or sticking to tasks.*	✔			✔			✔
28. *I withdraw from my family and friends.*	✔	✔					
29. *I act without thinking and don't worry about what will happen.*				✔	✔		
30. *I feel like I don't have any friends or that no one likes me.*	✔	✔					

Note: Child/Parent INT=Items on child/parent Internalizing factor; Child/Parent EXT=Items on child/parent Externalizing factor; No INT/EXT-Child/Parent=Items that did not load onto either factor for children/parents; Switched factors=Items that switched from INT to EXT or EXT to INT when comparing parents and children.

Table 4

YOQ Items by Factor

YOQ Item	Child INT T1	Child EXT T1	Parent INT T1	Parent EXT T1	No INT/EXT Child T1	No INT/EXT Parent T1	Switched Parent and Child T1	Child INT T2	Child EXT T2	Parent INT T2	Parent EXT T2	No INT/EXT Child T2	No INT/EXT Parent T2	Switched Parent and Child T2	Consistent Across P/C	Consistent Across T1 &T2
1. I have headaches or feel dizzy.	✓		✓					✓		✓					✓	✓
2. I don't participate in activities that used to be fun.	✓					✓		✓		✓						
3. I argue or speak rudely to others.		✓		✓					✓		✓				✓	✓
4. I have a hard time finishing my assignments or I do them carelessly.	✓			✓			✓	✓			✓			✓		✓
5. My emotions are strong and change quickly.	✓					✓		✓		✓						
6. I have physical fights (hitting, kicking, biting, or scratching) with my family or others my age.		✓		✓					✓		✓				✓	✓
7. I worry and can't get thoughts out of my mind.	✓		✓					✓		✓					✓	✓
8. I steal or lie.		✓		✓					✓		✓				✓	✓
11. I am tense and easily startled (jumpy).	✓		✓					✓		✓					✓	✓
12. I am sad or unhappy.	✓		✓					✓		✓					✓	✓
13. I have a hard time trusting friends, family members, or other adults.	✓		✓					✓		✓					✓	✓
14. I think that others are trying to hurt me even when they're not.	✓		✓							✓		✓				
16. I physically fight with adults.		✓		✓					✓		✓				✓	✓
17. My stomach hurts or I feel sick more than others my same age.	✓		✓							✓		✓				
18. I don't have friends or I don't keep friends very long.	✓					✓		✓		✓						
19. I think about suicide or feel I would be better off dead.	✓		✓					✓					✓			
20. I have nightmares, trouble getting to sleep, oversleeping, or waking up too early.	✓					✓		✓		✓						
21. I complain about or question rules, expectations, or responsibilities.				✓	✓				✓		✓					
22. I break rules, laws, or don't meet others' expectationss on purpose.		✓		✓					✓		✓				✓	✓
23. I feel irritated.	✓					✓		✓					✓		✓	✓
24. I get angry enough to threaten others.		✓		✓					✓				✓			
25. I get into trouble when I'm bored.		✓		✓					✓		✓				✓	✓
26. I destroy property on purpose.		✓		✓							✓	✓				
27. I have a hard time concentrating, thinking clearly, or sticking to tasks.	✓			✓			✓	✓			✓			✓		✓
28. I withdraw from my family and friends.	✓		✓					✓		✓					✓	✓
29. I act without thinking and don't worry about what will happen.		✓		✓					✓		✓				✓	✓
30. I feel like I don't have any friends or that no one likes me.	✓		✓					✓		✓					✓	✓
Total:	17	9	10	12	1	5	2	15	9	13	11	3	3	2	15	14

Note: Child/Parent INT T1/T2=Items on child/parent Internalizing factor at Time 1/Time2; Child/Parent EXT TI/T2=Items on child/parent Externalizing factor at Time1/Time2; No INT/EXT TI/T2-Child/Parent=Items that did not load onto either factor for children/parents; Switched factors T1/T2=Items that switched from INT to EXT or EXT to INT when comparing parents and children; Consistent across P/C=items that loaded onto the same parent/child INT/EXT factors; Consistent Across T1 & T2=items that loaded onto the same factors at Time 1 and Time 2

Table 5

Pearson Correlations-Children and Parents Time 1

	1	2	3
YOQ Total	-	.88**	.67**
INT_Total	.84**	-	.30**
EXT_Total	.86**	.50**	-

Note: Correlations above dotted line=Children T1; Correlations below dotted line=Parents T1

*Correlation is significant at the .05 level * *Correlation is significant at the .01 level

Table 6

Pearson Correlations-Children and Parents Time 2

	1	2	3
YOQ Total	-	.92**	.65**
INT_Total	.87**	-	.36**
EXT_Total	.84**	.51**	-

Note: Correlations above dotted line=Children T1; Correlations below dotted line=Parents T1

*Correlation is significant at the .05 level * *Correlation is significant at the .01 level

Table 7

Pearson Correlations-Children Time 1 and Children Time 2

	1	2	3
YOQ Total	-	.88**	.67**
INT_Total	.92**	-	.30**
EXT_Total	.65**	.36**	-

Note: Correlations above dotted line=Children T1; Correlations below dotted line=Children T2

*Correlation is significant at the .05 level * *Correlation is significant at the .01 level

Table 8

Pearson Correlations-Parents Time 1 and Parents Time 2

	1	2	3
YOQ Total	-	.84**	.86**
INT_Total	.87**	-	.50**
EXT_Total	.84**	.51**	-

Note: Correlations above dotted line=Parents T1; Correlations below dotted line=Parents T2

*Correlation is significant at the .05 level * *Correlation is significant at the .01 level

Table 9

Exploratory Factor Analysis Fit Indices-Two-Factor Fixed (27 Items)

	KMO	*Bartlett's Test of Sphericity*	*df*	*p*	*Correlation between Factor 1 and Factor 2*
Child Time 1	0.79	1149.82	351	<.001	0.30
Parent Time 1	0.81	1197.89	351	<.001	0.56
Child Time 2	0.77	1318.95	351	<.001	0.32
Parent Time 2	0.79	1214.60	351	<.001	0.52

Table 10

Exploratory Factor Analysis-Two-Factor Fixed Total Variance Explained (27 Items)

		Initial Eigenvalues		
		Total	*% of Variance*	*Cumulative %*
Child Time 1				
	Factor 1	8.89	32.93	32.92
	Factor 2	3.32	12.29	45.22
Parent Time 1				
	Factor 1	10.72	39.69	39.69
	Factor 2	2.43	8.99	48.68
Child Time 2				
	Factor 1	9.43	34.93	34.93
	Factor 2	3.28	12.15	47.08
Parent Time 2				
	Factor 1	10.13	37.52	37.52
	Factor 2	2.64	9.76	47.28

Table 11

Exploratory Factor Analysis Fit Indices-30 Item YOQ

	KMO	*Bartlett's Test of Sphericity*	*df*	*p*
Child Time 1	0.79	1298.24	435	<.001
Parent Time 1	0.76	1368.14	435	<.001
Child Time 2	0.72	1483.71	435	<.001
Parent Time 2	0.76	1383.94	435	<.001

Table 12

Exploratory Factor Analysis- Total Variance Explained-30 Item YOQ

		Initial Eigenvalues		
		Total	*% of Variance*	*Cumulative %*
Child Time 1				
	Factor 1	9.38	31.26	31.26
	Factor 2	3.46	11.54	42.80
Parent Time 1				
	Factor 1	11.19	37.31	37.31
	Factor 2	2.58	8.61	45.92
Child Time 2				
	Factor 1	9.83	32.77	32.77
	Factor 2	3.46	11.54	44.30
Parent Time 2				
	Factor 1	10.69	35.63	35.63
	Factor 2	2.71	9.04	44.67

Table 13

Two-Factor Fixed Exploratory Factor Analysis-Child Time 1

	Initial	*Pattern Matrix*[2]	
Y-OQ-SR Item	*Communalities*[1]	*INT Loadings*	*EXT Loadings*
Y-OQ_12	.85	.83	.10
Y-OQ_7	.76	.80	-.15
Y-OQ_28	.83	.79	-.04
Y-OQ_30	.66	.71	.04
Y-OQ_27	.68	.68	-.03
Y-OQ_2	.65	.68	-.17
Y-OQ_13	.69	.68	-.03
Y-OQ_11	.64	.67	-.14
Y-OQ_19	.66	.64	.18
Y-OQ_5	.64	.63	.04
Y-OQ_4	.59	.60	.12
Y-OQ_14	.65	.58	.11
Y-OQ_20	.52	.54	-.01
Y-OQ_23	.62	.51	.23
Y-OQ_1	.55	.46	.02
Y-OQ_17	.55	.45	.12
Y-OQ_18	.52	.34	.05
Y-OQ_6	.67	-.19	.79
Y-OQ_22	.75	.19	.72
Y-OQ_24	.62	.18	.66
Y-OQ_16	.50	-.26	.59
Y-OQ_25	.53	-.08	.57
Y-OQ_3	.69	.21	.53
Y-OQ_8	.45	.21	.49
Y-OQ_26	.53	.10	.47
Y-OQ_29	.64	.21	.41
Y-OQ_21	.59	.23	.40

Note: [1]Extraction Method: Principal Axis Factoring; [2]Rotation Method: Promax with Kaiser Normalization

Table 14

Two-Factor Fixed Exploratory Factor Analysis-Child Time 2

	Initial	*Pattern Matrix*[2]	
Y-OQ-SR Item	*Communalities*[1]	*INT Loadings*	*EXT Loadings*
Y-OQ_12	.85	.91	-.09
Y-OQ_27	.86	.81	-.13
Y-OQ_23	.83	.8	.06
Y-OQ_28	.80	.78	-.03
Y-OQ_7	.76	.78	-.12
Y-OQ_30	.70	.74	-.09
Y-OQ_5	.73	.67	.08
Y-OQ_18	.76	.66	.06
Y-OQ_20	.68	.64	-.10
Y-OQ_2	.67	.62	.07
Y-OQ_13	.72	.61	.09
Y-OQ_11	.60	.60	.01
Y-OQ_1	.79	.53	.25
Y-OQ_4	.66	.48	.12
Y-OQ_19	.63	.43	.11
Y-OQ_14	.65	.35	.22
Y-OQ_25	.70	-.06	.80
Y-OQ_3	.80	.28	.68
Y-OQ_22	.83	.14	.66
Y-OQ_24	.74	.06	.63
Y-OQ_21	.70	.24	.59
Y-OQ_29	.80	.20	.58
Y-OQ_8	.57	-.07	.54
Y-OQ_6	.53	-.01	.50
Y-OQ_17	.80	.32	.39
Y-OQ_26	.69	.25	.33
Y-OQ_16	.49	-.19	.31

Note: [1]Extraction Method: Principal Axis Factoring; [2]Rotation Method: Promax with Kaiser Normalization

Table 15

Two-Factor Fixed Exploratory Factor Analysis-Parent Time 1

	Initial	*Pattern Matrix*[2]	
Y-OQ-PR Item	*Communalities*[1]	*EXT Loadings*	*INT Loadings*
Y-OQ_25	.85	.98	-.24
Y-OQ_29	.81	.85	-.10
Y-OQ_22	.77	.69	.13
Y-OQ_6	.66	.65	-.04
Y-OQ_3	.71	.65	.05
Y-OQ_26	.74	.63	.08
Y-OQ_8	.73	.63	-.07
Y-OQ_27	.71	.63	.11
Y-OQ_24	.76	.63	.25
Y-OQ_4	.69	.62	.15
Y-OQ_21	.71	.61	.09
Y-OQ_16	.67	.46	.16
Y-OQ_18	.66	.34	.29
Y-OQ_2	.69	.27	.24
Y-OQ_30	.74	-.04	.85
Y-OQ_28	.85	.11	.75
Y-OQ_14	.74	.01	.69
Y-OQ_12	.72	.14	.67
Y-OQ_13	.65	.09	.65
Y-OQ_19	.54	-.15	.60
Y-OQ_11	.56	.10	.57
Y-OQ_7	.58	-.11	.54
Y-OQ_23	.74	.32	.50
Y-OQ_5	.79	.42	.45
Y-OQ_17	.58	.07	.38
Y-OQ_20	.46	.24	.38
Y-OQ_1	.58	.06	.33

Note: [1]Extraction Method: Principal Axis Factoring; [2]Rotation Method: Promax with Kaiser Normalization

Table 16

Two-Factor Fixed Exploratory Factor Analysis-Parent Time 2

	Initial	*Pattern Matrix*[2]	
Y-OQ-PR Item	*Communalities*[1]	*INT Loadings*	*EXT Loadings*
Y-OQ_12	.85	.83	.00
Y-OQ_28	.74	.82	-.04
Y-OQ_7	.77	.80	-.02
Y-OQ_2	.75	.66	.02
Y-OQ_30	.77	.65	.00
Y-OQ_14	.70	.62	.08
Y-OQ_17	.74	.60	-.17
Y-OQ_1	.71	.54	-.07
Y-OQ_5	.78	.52	.23
Y-OQ_18	.83	.50	.27
Y-OQ_13	.78	.49	.24
Y-OQ_11	.66	.48	.21
Y-OQ_23	.83	.45	.34
Y-OQ_20	.69	.45	.20
Y-OQ_29	.86	-.04	.85
Y-OQ_21	.88	-.03	.83
Y-OQ_8	.82	-.10	.82
Y-OQ_25	.79	.00	.75
Y-OQ_22	.75	-.07	.75
Y-OQ_3	.67	.18	.58
Y-OQ_4	.81	.00	.54
Y-OQ_27	.82	.23	.51
Y-OQ_6	.75	.17	.46
Y-OQ_16	.73	.14	.46
Y-OQ_24	.75	.31	.36
Y-OQ_26	.71	-.04	.30
Y-OQ_19	.54	.20	.27

Note: [1]Extraction Method: Principal Axis Factoring; [2]Rotation Method: Promax with Kaiser Normalization

Table 17

Exploratory Factor Analysis for IRT (30 items)-Initial Communalities[1]

Y-OQ-Item	*Child T1*	*Child T2*	*Parent T1*	*Parent T2*
Y-OQ_1	.56	.83	.65	.76
Y-OQ_2	.65	.70	.70	.77
Y-OQ_3	.70	.81	.71	.69
Y-OQ_4	.64	.72	.70	.82
Y-OQ_5	.66	.78	.83	.78
Y-OQ_6	.71	.54	.75	.80
Y-OQ_7	.76	.77	.62	.79
Y-OQ_8	.49	.67	.73	.83
Y-OQ_9	.57	.63	.70	.65
Y-OQ_10	.60	.56	.71	.72
Y-OQ_11	.65	.67	.60	.74
Y-OQ_12	.86	.86	.80	.88
Y-OQ_13	.70	.73	.67	.80
Y-OQ_14	.67	.72	.79	.74
Y-OQ_15	.52	.71	.68	.80
Y-OQ_16	.53	.57	.75	.75
Y-OQ_17	.60	.81	.60	.75
Y-OQ_18	.56	.79	.77	.85
Y-OQ_19	.73	.72	.66	.81
Y-OQ_20	.57	.73	.52	.71
Y-OQ_21	.59	.73	.73	.90
Y-OQ_22	.75	.85	.80	.76
Y-OQ_23	.66	.85	.75	.87
Y-OQ_24	.67	.74	.77	.76
Y-OQ_25	.56	.71	.90	.85
Y-OQ_26	.53	.76	.74	.77
Y-OQ_27	.71	.88	.76	.83
Y-OQ_28	.85	.82	.86	.77
Y-OQ_29	.66	.81	.87	.88
Y-OQ_30	.67	.76	.75	.78

Note: [1]Extraction Method: Principal Axis Factoring

Table 18

Group Parameter Estimates-Child Time 1 vs. Parent Time 1

Group	μ	*s.e.*	σ^2	*s.e.*	σ	*s.e.*
Parent T1	0	-----	1	-----	1	-----
Child T1	0.05	0.18	0.86	0.36	0.93	0.2

Table 19

Group Parameter Estimates-Child Time 2 vs. Parent Time 2

Group	μ	*s.e.*	σ^2	*s.e.*	σ	*s.e.*
Parent T2	0	-----	1	-----	1	-----
Child T2	-0.32	0.27	1.52	0.66	1.23	0.27

Table 20

Group Parameter Estimates-Child Time 1 vs. Child Time 2

Group	μ	*s.e.*	σ^2	*s.e.*	σ	*s.e.*
Child T1	0	-----	1	-----	1	-----
Child T2	-0.84	0.19	1.87	0.46	1.37	0.17

Table 21

Group Parameter Estimates-Parent Time 1 vs. Parent Time 2

Group	μ	*s.e.*	σ^2	*s.e.*	σ	*s.e.*
Parent T1	0	-----	1	-----	1	-----
Parent T2	-0.42	0.16	1.03	0.28	1.02	0.14

Table 22

Likelihood-based Values and Goodness of Fit Statistics [1]

Comparison Group	*2loglikelihood*	*Akaike Information Criterion (AIC):*	*Bayesian Information Criterion (BIC):*
CT1 vs. PT1	9710.47	10026.47	10501.09
CT2 vs. PT2	7120.84	7394.84	7794.88
CT1 vs. CT2	8901.58	9171.58	9580.69
PT1 vs. PT2	7728.41	8010.41	8417.95

[1] Statistics based on the loglikelihood

Table 23

Graded Model Item Parameter Estimates for Parents, Parents vs. Children at Time 1[1]

Item	*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
YOQ_1	0.95	0.25	-0.43	0.29	0.75	0.24	3.54	0.87		
YOQ_2	1.27	0.29	-0.66	0.29	0.27	0.18	1.8	0.35	3.02	0.64
YOQ_3	1.25	0.29	-1.56	0.45	0	0.2	1.44	0.29	3.06	0.66
YOQ_4	1.79	0.39	-1.18	0.35	-0.28	0.21	0.84	0.17	2.24	0.41
YOQ_5	1.94	0.41	-1.04	0.33	-0.23	0.2	0.96	0.17	1.93	0.33
YOQ_6	0.85	0.26	0.55	0.25	1.9	0.51	3.9	1.11	4.61	1.36
YOQ_7	0.95	0.3	-0.83	0.42	0.28	0.29	1.72	0.53	4.1	1.31
YOQ_8	1.11	0.28	-0.22	0.24	0.87	0.23	2.18	0.48	4.38	1.14
YOQ_9	0.82	0.23	-0.29	0.29	1.18	0.34	3.19	0.85	4.72	1.34
YOQ_10	0.95	0.34	2.31	0.69	3.22	1.01	4.93	1.71		
YOQ_11	1.63	0.37	-0.12	0.2	1.02	0.19	1.98	0.36	2.84	0.57
YOQ_12	2.79	0.55	-0.73	0.26	-0.05	0.17	0.98	0.15	1.96	0.31
YOQ_13	1.73	0.38	-0.48	0.25	0.31	0.16	1.18	0.21	2.45	0.46
YOQ_14	2.07	0.47	0.44	0.15	1.23	0.2	2.05	0.36		
YOQ_15	1.63	0.4	0.86	0.18	1.78	0.33	2.47	0.5		
YOQ_16	1.85	0.59	1.26	0.29	1.71	0.39	2.36	0.57		
YOQ_17	1.13	0.28	0.31	0.2	1.51	0.33	2.72	0.61		
YOQ_18	1.41	0.35	0.55	0.17	1.6	0.32	2.38	0.5	2.68	0.57
YOQ_19	1.84	0.41	0.71	0.16	1.21	0.21	1.93	0.34	3.06	0.63
YOQ_20	1.26	0.29	-0.81	0.31	0.15	0.19	1.08	0.23	1.99	0.4
YOQ_21	1.12	0.27	-1.45	0.45	-0.19	0.23	1.15	0.27	2.44	0.53
YOQ_22	1.77	0.4	-0.13	0.2	1.02	0.19	1.84	0.33	2.93	0.61
YOQ_23	1.9	0.4	-0.77	0.28	0	0.17	1.21	0.21	2.19	0.39
YOQ_24	1.96	0.44	0.33	0.15	0.98	0.17	1.91	0.33	2.58	0.5
YOQ_25	1.05	0.28	0.15	0.21	1.37	0.32	2.7	0.64	3.88	0.99
YOQ_26	1.44	0.37	0.88	0.2	1.73	0.34	2.72	0.59	3.58	0.87
YOQ_27	1.7	0.37	-0.99	0.32	-0.14	0.19	0.91	0.18	2.3	0.43
YOQ_28	2.56	0.54	-0.2	0.19	0.29	0.14	1.44	0.22	1.99	0.33
YOQ_29	1.32	0.32	-0.48	0.26	0.49	0.18	1.28	0.27	2.37	0.5
YOQ_30	2.49	0.61	0.09	0.19	0.8	0.18	2.01	0.39		

[1] logit: $a(\theta - b)$

Table 24

Graded Model Item Parameter Estimates for Children, Parents vs. Children at Time 1[1]

Item	*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
YOQ_1	0.95	0.25	-0.43	0.29	0.75	0.24	3.54	0.87		
YOQ_2	1.27	0.29	-0.66	0.29	0.27	0.18	1.8	0.35	3.02	0.64
YOQ_3	1.25	0.29	-1.56	0.45	0	0.2	1.44	0.29	3.06	0.66
YOQ_4	1.79	0.39	-1.18	0.35	-0.28	0.21	0.84	0.17	2.24	0.41
YOQ_5	1.94	0.41	-1.04	0.33	-0.23	0.2	0.96	0.17	1.93	0.33
YOQ_6	0.85	0.26	0.55	0.25	1.9	0.51	3.9	1.11	4.61	1.36
YOQ_7	1.95	0.52	-0.74	0.33	-0.08	0.21	0.66	0.19	1.37	0.29
YOQ_8	1.11	0.28	-0.22	0.24	0.87	0.23	2.18	0.48	4.38	1.14
YOQ_9	0.82	0.23	-0.29	0.29	1.18	0.34	3.19	0.85	4.72	1.34
YOQ_10	0.95	0.34	2.31	0.69	3.22	1.01	4.93	1.71		
YOQ_11	1.63	0.37	-0.12	0.2	1.02	0.19	1.98	0.36	2.84	0.57
YOQ_12	2.79	0.55	-0.73	0.26	-0.05	0.17	0.98	0.15	1.96	0.31
YOQ_13	1.73	0.38	-0.48	0.25	0.31	0.16	1.18	0.21	2.45	0.46
YOQ_14	2.07	0.47	0.44	0.15	1.23	0.2	2.05	0.36		
YOQ_15	1.63	0.4	0.86	0.18	1.78	0.33	2.47	0.5		
YOQ_16	0.29	0.33	4.92	5.48	10.42	11.78	15.44	17.71		
YOQ_17	1.13	0.28	0.31	0.2	1.51	0.33	2.72	0.61		
YOQ_18	1.41	0.35	0.55	0.17	1.6	0.32	2.38	0.5	2.68	0.57
YOQ_19	1.84	0.41	0.71	0.16	1.21	0.21	1.93	0.34	3.06	0.63
YOQ_20	1.26	0.29	-0.81	0.31	0.15	0.19	1.08	0.23	1.99	0.4
YOQ_21	1.12	0.27	-1.45	0.45	-0.19	0.23	1.15	0.27	2.44	0.53
YOQ_22	1.77	0.4	-0.13	0.2	1.02	0.19	1.84	0.33	2.93	0.61
YOQ_23	1.9	0.4	-0.77	0.28	0	0.17	1.21	0.21	2.19	0.39
YOQ_24	1.96	0.44	0.33	0.15	0.98	0.17	1.91	0.33	2.58	0.5
YOQ_25	1.05	0.28	0.15	0.21	1.37	0.32	2.7	0.64	3.88	0.99
YOQ_26	1.44	0.37	0.88	0.2	1.73	0.34	2.72	0.59	3.58	0.87
YOQ_27	1.7	0.37	-0.99	0.32	-0.14	0.19	0.91	0.18	2.3	0.43
YOQ_28	2.56	0.54	-0.2	0.19	0.29	0.14	1.44	0.22	1.99	0.33
YOQ_29	1.32	0.32	-0.48	0.26	0.49	0.18	1.28	0.27	2.37	0.5
YOQ_30	3.36	0.96	0.59	0.15	1	0.18	1.34	0.23		

[1] logit: $a(\theta - b)$

Table 25

DIF Statistics for Graded Items-Child Time 1 vs. Parent Time 1

Parents	*Children*	*Total* X_2	*d.f.*	*p*	X_{2a}	*d.f.*	*p*	$X_{2c\|a}$	*d.f.*	*p*
1	1	0	4	1	0	1	1	0	3	1
2	2	0	5	1	0	1	1	0	4	1
3	3	0	5	1	0	1	1	0	4	1
4	4	0	5	1	0	1	1	0	4	1
5	5	0	5	1	0	1	1	0	4	1
6	6	0	5	1	0	1	1	0	4	1
7	7	9.5	5	0.09	2.8	1	0.10	6.7	4	0.15
8	8	0	5	1	0	1	1	0	4	1
9	9	0	5	1	0	1	1	0	4	1
10	10	0	4	1	0	1	1	0	3	1
11	11	0	5	1	0	1	1	0	4	1
12	12	0	5	1	0	1	1	0	4	1
13	13	0	5	1	0	1	1	0	4	1
14	14	0	4	1	0	1	1	0	3	1
15	15	0	4	1	0	1	1	0	3	1
16	16	8.1	4	0.09	5.4	1	0.02	2.6	3	0.45
17	17	0	4	1	0	1	1	0	3	1
18	18	0	5	1	0	1	1	0	4	1
19	19	0	5	1	0	1	1	0	4	1
20	20	0	5	1	0	1	1	0	4	1
21	21	0	5	1	0	1	1	0	4	1
22	22	0	5	1	0	1	1	0	4	1
23	23	0	5	1	0	1	1	0	4	1
24	24	0	5	1	0	1	1	0	4	1
25	25	0	5	1	0	1	1	0	4	1
26	26	0	5	1	0	1	1	0	4	1
27	27	0	5	1	0	1	1	0	4	1
28	28	0	5	1	0	1	1	0	4	1
29	29	0	5	1	0	1	1	0	4	1
30	30	8.5	4	0.07	0.6	1	0.45	7.9	3	0.05

Table 26

Graded Model Item Parameter Estimates for Parents, Parents vs. Children at Time 2[1]

Item	*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
YOQ_1	1.26	0.31	-0.02	0.24	1.24	0.27	2.99	0.68		
YOQ_2	1.73	0.4	-0.06	0.22	0.83	0.19	1.9	0.35	3.03	0.66
YOQ_3	1.62	0.38	-0.57	0.29	0.33	0.19	1.73	0.32		
YOQ_4	1.11	0.27	-1.13	0.41	-0.01	0.24	1.56	0.34	2.45	0.54
YOQ_5	1.8	0.38	-0.69	0.3	0.31	0.18	1.5	0.26	2.28	0.41
YOQ_6	1.86	0.59	0.68	0.22	1.15	0.28				
YOQ_7	1.74	0.39	-0.51	0.28	0.11	0.2	1.5	0.26	2.64	0.53
YOQ_8	1.1	0.31	0.58	0.23	1.53	0.37	3.3	0.85		
YOQ_9	1.17	0.29	-0.56	0.31	0.79	0.23	2.05	0.44	3.18	0.74
YOQ_10	0.66	0.38	4.01	2.09	5.95	3.24				
YOQ_11	1.86	0.46	0.31	0.18	1.1	0.21				
YOQ_12	2.57	0.57	-0.34	0.24	0.43	0.16	1.47	0.23		
YOQ_13	1.9	0.44	0	0.21	0.82	0.18	1.69	0.3	2.68	0.54
YOQ_14	1.85	0.5	1.02	0.2	1.77	0.34	2.52	0.54		
YOQ_15	1.4	0.46	1.66	0.39	2.41	0.6				
YOQ_16	1.01	0.39	2.18	0.68	3.82	1.32				
YOQ_17	1.32	0.35	0.68	0.21	1.6	0.34	2.91	0.68		
YOQ_18	2.35	0.54	0.38	0.17	1.02	0.18	1.94	0.33	2.64	0.54
YOQ_19	1.34	0.41	1.36	0.31	2.11	0.5				
YOQ_20	1.63	0.47	0.18	0.23	1.03	0.27	1.3	0.32	3.22	0.92
YOQ_21	1.51	0.34	-0.75	0.31	0.21	0.2	1.55	0.29	2.63	0.53
YOQ_22	1.41	0.45	0.14	0.25	1.01	0.31				
YOQ_23	2.4	0.51	-0.4	0.24	0.32	0.17	1.41	0.22	2.62	0.52
YOQ_24	1.68	0.45	0.92	0.2	1.91	0.38	2.84	0.64		
YOQ_25	1.36	0.36	0.45	0.2	1.52	0.33	3.03	0.72		
YOQ_26	1.09	0.39	2	0.56	3.22	0.99				
YOQ_27	1.81	0.39	-0.4	0.25	0.26	0.18	1.09	0.2	2.38	0.44
YOQ_28	2.28	0.51	0.11	0.19	1.01	0.18	2.07	0.37		
YOQ_29	1.5	0.37	-0.02	0.22	1	0.22	2.21	0.44	3.35	0.8
YOQ_30	1.86	0.43	0.28	0.18	1.06	0.2	1.95	0.35		

[1] logit: $a(\theta - b)$

Table 27

Graded Model Item Parameter Estimates for Children, Parents vs. Children at Time 2[1]

Item	*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
YOQ_1	1.26	0.31	-0.02	0.24	1.24	0.27	2.99	0.68		
YOQ_2	1.73	0.4	-0.06	0.22	0.83	0.19	1.9	0.35	3.03	0.66
YOQ_3	1.62	0.38	-0.57	0.29	0.33	0.19	1.73	0.32		
YOQ_4	1.11	0.27	-1.13	0.41	-0.01	0.24	1.56	0.34	2.45	0.54
YOQ_5	1.8	0.38	-0.69	0.3	0.31	0.18	1.5	0.26	2.28	0.41
YOQ_6	0.85	0.42	1.86	0.81	5.22	2.57				
YOQ_7	1.74	0.39	-0.51	0.28	0.11	0.2	1.5	0.26	2.64	0.53
YOQ_8	1.1	0.31	0.58	0.23	1.53	0.37	3.3	0.85		
YOQ_9	1.17	0.29	-0.56	0.31	0.79	0.23	2.05	0.44	3.18	0.74
YOQ_10	0.66	0.38	4.01	2.09	5.95	3.24				
YOQ_11	1.86	0.46	0.31	0.18	1.1	0.21				
YOQ_12	2.57	0.57	-0.34	0.24	0.43	0.16	1.47	0.23		
YOQ_13	1.9	0.44	0	0.21	0.82	0.18	1.69	0.3	2.68	0.54
YOQ_14	1.85	0.5	1.02	0.2	1.77	0.34	2.52	0.54		
YOQ_15	1.4	0.46	1.66	0.39	2.41	0.6				
YOQ_16	1.01	0.39	2.18	0.68	3.82	1.32				
YOQ_17	1.32	0.35	0.68	0.21	1.6	0.34	2.91	0.68		
YOQ_18	2.35	0.54	0.38	0.17	1.02	0.18	1.94	0.33	2.64	0.54
YOQ_19	1.34	0.41	1.36	0.31	2.11	0.5				
YOQ_20	1.3	0.37	-0.75	0.38	0.28	0.25	1.41	0.38	2.02	0.53
YOQ_21	1.51	0.34	-0.75	0.31	0.21	0.2	1.55	0.29	2.63	0.53
YOQ_22	1.5	0.54	0.88	0.28	2.11	0.6				
YOQ_23	2.4	0.51	-0.4	0.24	0.32	0.17	1.41	0.22	2.62	0.52
YOQ_24	1.68	0.45	0.92	0.2	1.91	0.38	2.84	0.64		
YOQ_25	1.36	0.36	0.45	0.2	1.52	0.33	3.03	0.72		
YOQ_26	1.09	0.39	2	0.56	3.22	0.99				
YOQ_27	1.81	0.39	-0.4	0.25	0.26	0.18	1.09	0.2	2.38	0.44
YOQ_28	2.28	0.51	0.11	0.19	1.01	0.18	2.07	0.37		
YOQ_29	1.5	0.37	-0.02	0.22	1	0.22	2.21	0.44	3.35	0.8
YOQ_30	1.86	0.43	0.28	0.18	1.06	0.2	1.95	0.35		

[1] logit: $a(\theta - b)$

Table 28

DIF Statistics for Graded Items-Child Time 2 vs. Parent Time 2

Parents	*Children*	*Total* X_2	*d.f.*	*p*	X_{2a}	*d.f.*	*p*	$X_{2c\|a}$	*d.f.*	*p*
1	1	0	4	1	0	1	1	0	3	1
2	2	0	5	1	0	1	1	0	4	1
3	3	0	4	1	0	1	1	0	3	1
4	4	0	5	1	0	1	1	0	4	1
5	5	0	5	1	0	1	1	0	4	1
6	6	9.4	3	0.02	1.9	1	0.16	7.4	2	0.02
7	7	0	5	1	0	1	1	0	4	1
8	8	0	4	1	0	1	1	0	3	1
9	9	0	5	1	0	1	1	0	4	1
10	10	0	3	1	0	1	1	0	2	1
11	11	0	3	1	0	1	1	0	2	1
12	12	0	4	1	0	1	1	0	3	1
13	13	0	5	1	0	1	1	0	4	1
14	14	0	4	1	0	1	1	0	3	1
15	15	0	3	1	0	1	1	0	2	1
16	16	0	3	1	0	1	1	0	2	1
17	17	0	4	1	0	1	1	0	3	1
18	18	0	5	1	0	1	1	0	4	1
19	19	0	3	1	0	1	1	0	2	1
20	20	15.7	5	0.01	0.3	1	0.58	15.4	4	0.00
21	21	0	5	1	0	1	1	0	4	1
22	22	7.7	3	0.05	0	1	0.89	7.7	2	0.02
23	23	0	5	1	0	1	1	0	4	1
24	24	0	4	1	0	1	1	0	3	1
25	25	0	4	1	0	1	1	0	3	1
26	26	0	3	1	0	1	1	0	2	1
27	27	0	5	1	0	1	1	0	4	1
28	28	0	4	1	0	1	1	0	3	1
29	29	0	5	1	0	1	1	0	4	1
30	30	0	4	1	0	1	1	0	3	1

Table 29

Graded Model Item Parameter Estimates for Time 1, Children Time 1 vs. Children Time 2[1]

Item	*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
YOQ_1	1.25	0.22	-0.7	0.2	0.45	0.18	2.66	0.47		
YOQ_2	1.41	0.23	-0.56	0.17	0.39	0.16	1.56	0.27	3.41	0.68
YOQ_3	1.21	0.2	-1.46	0.28	0.07	0.17	1.8	0.31		
YOQ_4	1.26	0.21	-1.58	0.28	-0.38	0.17	1.35	0.25	2.48	0.43
YOQ_5	1.69	0.25	-1.07	0.18	-0.08	0.13	1	0.17	1.78	0.27
YOQ_6	0.46	0.24	0.27	0.51	2.44	1.3				
YOQ_7	1.79	0.28	-0.94	0.18	-0.27	0.13	0.73	0.15	1.55	0.23
YOQ_8	0.93	0.19	-0.2	0.2	1.27	0.32	2.86	0.62		
YOQ_9	0.78	0.16	-1.01	0.29	0.93	0.31	2.94	0.68	5.07	1.25
YOQ_10	0.94	0.29	2.02	0.58	3.16	0.93				
YOQ_11	1.67	0.27	-0.21	0.14	1.03	0.18	1.91	0.3		
YOQ_12	3.23	0.47	-0.72	0.12	-0.06	0.1	0.82	0.11	1.63	0.18
YOQ_13	1.67	0.25	-0.63	0.15	0.2	0.13	1.24	0.2	2.29	0.36
YOQ_14	1.59	0.28	0.32	0.14	1.31	0.22	2.25	0.38		
YOQ_15	1.4	0.29	0.86	0.19	1.76	0.32				
YOQ_16	0.29	0.21	5.81	4.36	11.65	8.65				
YOQ_17	1.28	0.24	0.36	0.17	1.34	0.26	2.35	0.44		
YOQ_18	1.35	0.25	0.47	0.16	1.72	0.31	2.64	0.5	3.2	0.65
YOQ_19	2.37	0.41	0.62	0.12	1.05	0.15	1.7	0.23		
YOQ_20	1.19	0.2	-1.22	0.24	-0.02	0.17	0.93	0.21	1.86	0.34
YOQ_21	1.08	0.19	-1.42	0.27	0.08	0.18	1.65	0.33	2.83	0.54
YOQ_22	1.46	0.27	0.13	0.15	1.56	0.27				
YOQ_23	1.98	0.3	-0.81	0.15	-0.06	0.12	1.03	0.16	2.23	0.33
YOQ_24	1.3	0.25	0.42	0.16	1.41	0.27	2.47	0.46	3.64	0.79
YOQ_25	0.7	0.18	0.28	0.27	1.9	0.55	4.13	1.12		
YOQ_26	1.14	0.29	1.51	0.36	2.7	0.65				
YOQ_27	1.69	0.26	-0.91	0.17	-0.05	0.13	0.96	0.17	2.18	0.34
YOQ_28	2.37	0.35	-0.33	0.12	0.27	0.11	1.41	0.18	2.03	0.28
YOQ_29	1.07	0.2	-0.66	0.2	0.71	0.22	1.74	0.35	2.9	0.57
YOQ_30	2.03	0.33	0.28	0.12	0.94	0.15	1.47	0.2	2.51	0.4

[1] logit: $a(\theta - b)$

Table 30

Graded Model Item Parameter Estimates for Time 2, Children Time 1 vs. Children Time 2[1]

Item	*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
YOQ_1	1.25	0.22	-0.7	0.2	0.45	0.18	2.66	0.47		
YOQ_2	1.41	0.23	-0.56	0.17	0.39	0.16	1.56	0.27	3.41	0.68
YOQ_3	1.21	0.2	-1.46	0.28	0.07	0.17	1.8	0.31		
YOQ_4	1.26	0.21	-1.58	0.28	-0.38	0.17	1.35	0.25	2.48	0.43
YOQ_5	1.69	0.25	-1.07	0.18	-0.08	0.13	1	0.17	1.78	0.27
YOQ_6	0.71	0.33	1.72	0.95	5.71	3				
YOQ_7	1.79	0.28	-0.94	0.18	-0.27	0.13	0.73	0.15	1.55	0.23
YOQ_8	0.93	0.19	-0.2	0.2	1.27	0.32	2.86	0.62		
YOQ_9	0.78	0.16	-1.01	0.29	0.93	0.31	2.94	0.68	5.07	1.25
YOQ_10	0.94	0.29	2.02	0.58	3.16	0.93				
YOQ_11	1.67	0.27	-0.21	0.14	1.03	0.18	1.91	0.3		
YOQ_12	3.23	0.47	-0.72	0.12	-0.06	0.1	0.82	0.11	1.63	0.18
YOQ_13	1.67	0.25	-0.63	0.15	0.2	0.13	1.24	0.2	2.29	0.36
YOQ_14	1.59	0.28	0.32	0.14	1.31	0.22	2.25	0.38		
YOQ_15	1.4	0.29	0.86	0.19	1.76	0.32				
YOQ_16	0.29	0.21	5.81	4.36	11.65	8.65				
YOQ_17	1.28	0.24	0.36	0.17	1.34	0.26	2.35	0.44		
YOQ_18	1.35	0.25	0.47	0.16	1.72	0.31	2.64	0.5	3.2	0.65
YOQ_19	2.37	0.41	0.62	0.12	1.05	0.15	1.7	0.23		
YOQ_20	1.19	0.2	-1.22	0.24	-0.02	0.17	0.93	0.21	1.86	0.34
YOQ_21	1.08	0.19	-1.42	0.27	0.08	0.18	1.65	0.33	2.83	0.54
YOQ_22	1.46	0.27	0.13	0.15	1.56	0.27				
YOQ_23	1.98	0.3	-0.81	0.15	-0.06	0.12	1.03	0.16	2.23	0.33
YOQ_24	1.3	0.25	0.42	0.16	1.41	0.27	2.47	0.46	3.64	0.79
YOQ_25	0.7	0.18	0.28	0.27	1.9	0.55	4.13	1.12		
YOQ_26	1.14	0.29	1.51	0.36	2.7	0.65				
YOQ_27	1.69	0.26	-0.91	0.17	-0.05	0.13	0.96	0.17	2.18	0.34
YOQ_28	2.37	0.35	-0.33	0.12	0.27	0.11	1.41	0.18	2.03	0.28
YOQ_29	1.07	0.2	-0.66	0.2	0.71	0.22	1.74	0.35	2.9	0.57
YOQ_30	2.03	0.33	0.28	0.12	0.94	0.15	1.47	0.2	2.51	0.4

[1] logit: $a(\theta - b)$

Table 31

DIF Statistics for Graded Items-Child Time 1 vs. Child Time 2

Time 1	*Time 2*	*Total* X_2	*d.f.*	*p*	X_{2a}	*d.f.*	*p*	$X_{2c\|a}$	*d.f.*	*p*
1	1	0	4	1	0	1	1	0	3	1
2	2	0	5	1	0	1	1	0	4	1
3	3	0	4	1	0	1	1	0	3	1
4	4	0	5	1	0	1	1	0	4	1
5	5	0	5	1	0	1	1	0	4	1
6	6	12.6	3	0.01	0.4	1	0.54	12.2	2	0.00
7	7	0	5	1	0	1	1	0	4	1
8	8	0	4	1	0	1	1	0	3	1
9	9	0	5	1	0	1	1	0	4	1
10	10	0	3	1	0	1	1	0	2	1
11	11	0	4	1	0	1	1	0	3	1
12	12	0	5	1	0	1	1	0	4	1
13	13	0	5	1	0	1	1	0	4	1
14	14	0	4	1	0	1	1	0	3	1
15	15	0	3	1	0	1	1	0	2	1
16	16	0	3	1	0	1	1	0	2	1
17	17	0	4	1	0	1	1	0	3	1
18	18	0	5	1	0	1	1	0	4	1
19	19	0	4	1	0	1	1	0	3	1
20	20	0	5	1	0	1	1	0	4	1
21	21	0	5	1	0	1	1	0	4	1
22	22	0	3	1	0	1	1	0	2	1
23	23	0	5	1	0	1	1	0	4	1
24	24	0	5	1	0	1	1	0	4	1
25	25	0	4	1	0	1	1	0	3	1
26	26	0	3	1	0	1	1	0	2	1
27	27	0	5	1	0	1	1	0	4	1
28	28	0	5	1	0	1	1	0	4	1
29	29	0	5	1	0	1	1	0	4	1
30	30	0	5	1	0	1	1	0	4	1

Table 32

Graded Model Item Parameter Estimates for Time 1, Parents Time 1 vs. Parents Time 2[1]

Item	*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
YOQ_1	0.79	0.2	-0.12	0.24	1.44	0.43	4.21	1.13		
YOQ_2	1.37	0.25	-0.62	0.18	0.31	0.16	1.9	0.35	2.64	0.49
YOQ_3	1.71	0.28	-1.07	0.2	-0.15	0.14	1.01	0.19	2.82	0.5
YOQ_4	1.45	0.25	-1.35	0.26	-0.45	0.17	0.64	0.17	2.02	0.35
YOQ_5	1.98	0.31	-1.2	0.2	-0.33	0.13	1.03	0.17	2.25	0.36
YOQ_6	1.47	0.29	0.54	0.17	1.36	0.26	2.51	0.47		
YOQ_7	1.16	0.22	-0.88	0.24	0.02	0.18	1.58	0.33	3.57	0.76
YOQ_8	1.45	0.28	0.04	0.15	0.65	0.18	1.87	0.34		
YOQ_9	1.06	0.3	-0.01	0.25	1.06	0.35	3.12	0.86	3.41	0.96
YOQ_10	0.57	0.35	4.66	2.79	6.19	3.77				
YOQ_11	1.72	0.32	-0.05	0.14	0.72	0.17				
YOQ_12	2.09	0.34	-0.89	0.17	-0.02	0.12	1.36	0.21		
YOQ_13	1.69	0.28	-0.36	0.15	0.52	0.15	1.28	0.22	2.82	0.51
YOQ_14	2.41	0.46	0.68	0.13	1.3	0.19	2.03	0.31		
YOQ_15	1.81	0.4	1.09	0.2	1.89	0.34				
YOQ_16	1.99	0.43	1.1	0.19	1.72	0.29	2.19	0.38		
YOQ_17	0.97	0.23	0.22	0.21	1.58	0.39	3.37	0.81		
YOQ_18	1.71	0.3	0.13	0.14	0.83	0.17	1.82	0.3	2.29	0.4
YOQ_19	1.14	0.28	0.97	0.26	1.8	0.42				
YOQ_20	1.28	0.31	-0.61	0.26	0.12	0.21	1.4	0.35	2.15	0.51
YOQ_21	1.6	0.27	-1.2	0.22	-0.46	0.16	0.73	0.17	1.89	0.32
YOQ_22	1.77	0.31	-0.26	0.14	0.62	0.15	1.58	0.27		
YOQ_23	1.99	0.32	-0.88	0.17	-0.05	0.12	1.33	0.21	2.24	0.36
YOQ_24	2.49	0.43	0.36	0.11	1	0.15	1.96	0.29		
YOQ_25	1.9	0.33	0	0.13	0.95	0.17	1.91	0.31		
YOQ_26	1.44	0.31	0.78	0.19	1.64	0.32	2.71	0.55		
YOQ_27	1.85	0.3	-0.97	0.18	-0.3	0.13	0.65	0.15	2.17	0.35
YOQ_28	2.07	0.35	-0.25	0.13	0.56	0.13	1.87	0.29		
YOQ_29	1.95	0.32	-0.34	0.14	0.34	0.13	1.17	0.19	2.14	0.35
YOQ_30	1.67	0.3	-0.02	0.14	0.81	0.17	2.26	0.39		

[1] logit: $a(\theta - b)$

Table 33

Graded Model Item Parameter Estimates for Time 2, Parents Time 1 vs. Parents Time 2[1]

Item	*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
YOQ_1	0.79	0.2	-0.12	0.24	1.44	0.43	4.21	1.13		
YOQ_2	1.37	0.25	-0.62	0.18	0.31	0.16	1.9	0.35	2.64	0.49
YOQ_3	1.71	0.28	-1.07	0.2	-0.15	0.14	1.01	0.19	2.82	0.5
YOQ_4	1.45	0.25	-1.35	0.26	-0.45	0.17	0.64	0.17	2.02	0.35
YOQ_5	1.98	0.31	-1.2	0.2	-0.33	0.13	1.03	0.17	2.25	0.36
YOQ_6	1.47	0.29	0.54	0.17	1.36	0.26	2.51	0.47		
YOQ_7	1.16	0.22	-0.88	0.24	0.02	0.18	1.58	0.33	3.57	0.76
YOQ_8	1.45	0.28	0.04	0.15	0.65	0.18	1.87	0.34		
YOQ_9	1.32	0.35	-0.76	0.26	0.25	0.25	1.17	0.4	2.38	0.72
YOQ_10	0.57	0.35	4.66	2.79	6.19	3.77				
YOQ_11	1.72	0.32	-0.05	0.14	0.72	0.17				
YOQ_12	2.09	0.34	-0.89	0.17	-0.02	0.12	1.36	0.21		
YOQ_13	1.69	0.28	-0.36	0.15	0.52	0.15	1.28	0.22	2.82	0.51
YOQ_14	2.41	0.46	0.68	0.13	1.3	0.19	2.03	0.31		
YOQ_15	1.81	0.4	1.09	0.2	1.89	0.34				
YOQ_16	1.99	0.43	1.1	0.19	1.72	0.29	2.19	0.38		
YOQ_17	0.97	0.23	0.22	0.21	1.58	0.39	3.37	0.81		
YOQ_18	1.71	0.3	0.13	0.14	0.83	0.17	1.82	0.3	2.29	0.4
YOQ_19	1.14	0.28	0.97	0.26	1.8	0.42				
YOQ_20	1.55	0.41	-0.24	0.2	0.65	0.27	0.94	0.32	2.92	0.92
YOQ_21	1.6	0.27	-1.2	0.22	-0.46	0.16	0.73	0.17	1.89	0.32
YOQ_22	1.77	0.31	-0.26	0.14	0.62	0.15	1.58	0.27		
YOQ_23	1.99	0.32	-0.88	0.17	-0.05	0.12	1.33	0.21	2.24	0.36
YOQ_24	2.49	0.43	0.36	0.11	1	0.15	1.96	0.29		
YOQ_25	1.9	0.33	0	0.13	0.95	0.17	1.91	0.31		
YOQ_26	1.44	0.31	0.78	0.19	1.64	0.32	2.71	0.55		
YOQ_27	1.85	0.3	-0.97	0.18	-0.3	0.13	0.65	0.15	2.17	0.35
YOQ_28	2.07	0.35	-0.25	0.13	0.56	0.13	1.87	0.29		
YOQ_29	1.95	0.32	-0.34	0.14	0.34	0.13	1.17	0.19	2.14	0.35
YOQ_30	1.67	0.3	-0.02	0.14	0.81	0.17	2.26	0.39		

[1] logit: $a(\theta - b)$

Table 34

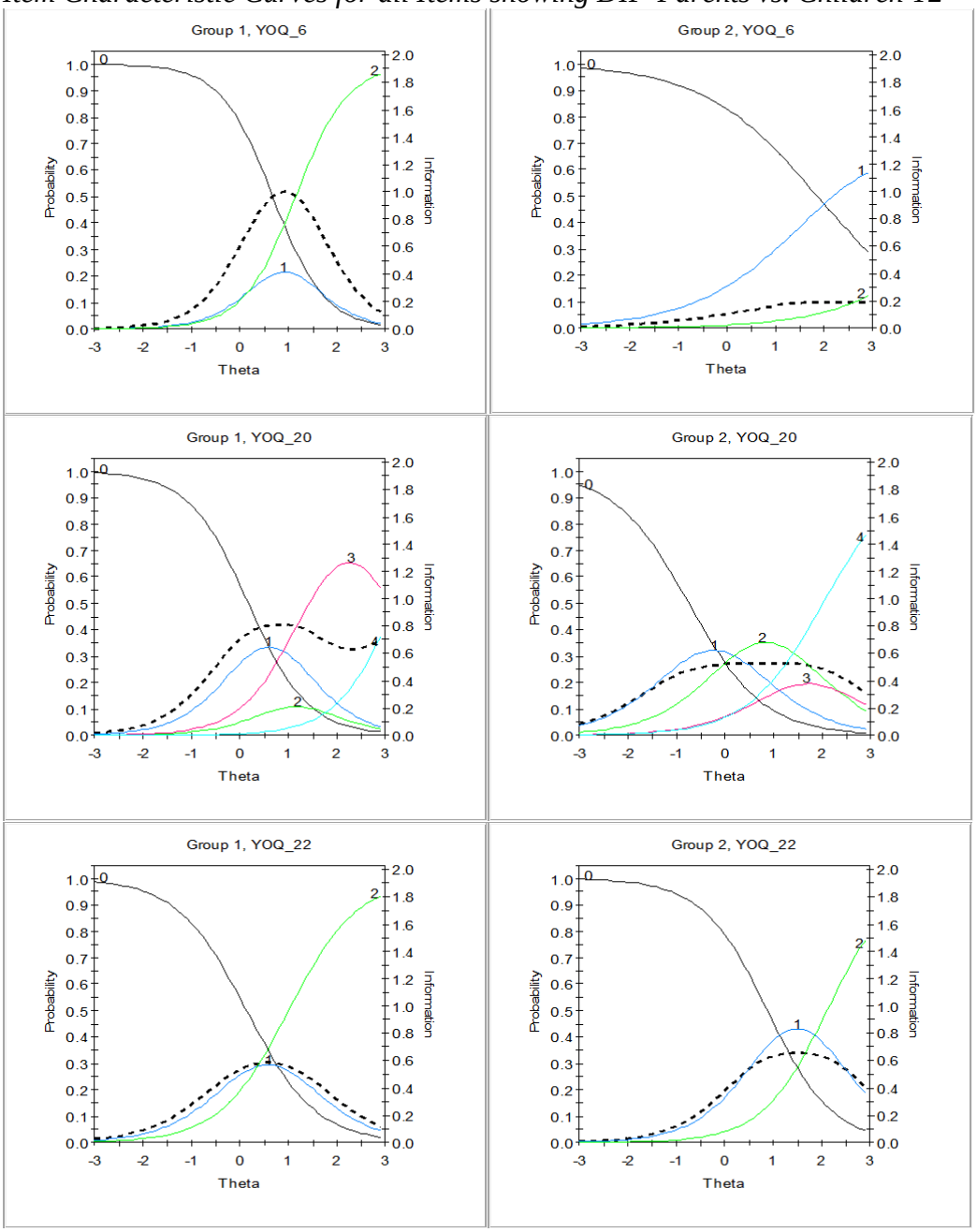
DIF Statistics for Graded Items-Parents Time 1 vs. Parents Time 2

Time 1	*Time 2*	*Total* X_2	*d.f.*	*p*	X_{2a}	*d.f.*	*p*	$X_{2c\|a}$	*d.f.*	*p*
1	1	0	4	1	0	1	1	0	3	1
2	2	0	5	1	0	1	1	0	4	1
3	3	0	5	1	0	1	1	0	4	1
4	4	0	5	1	0	1	1	0	4	1
5	5	0	5	1	0	1	1	0	4	1
6	6	0	4	1	0	1	1	0	3	1
7	7	0	5	1	0	1	1	0	4	1
8	8	0	4	1	0	1	1	0	3	1
9	9	13.2	5	0.02	0.3	1	0.59	12.9	4	0.01
10	10	0	3	1	0	1	1	0	2	1
11	11	0	3	1	0	1	1	0	2	1
12	12	0	4	1	0	1	1	0	3	1
13	13	0	5	1	0	1	1	0	4	1
14	14	0	4	1	0	1	1	0	3	1
15	15	0	3	1	0	1	1	0	2	1
16	16	0	4	1	0	1	1	0	3	1
17	17	0	4	1	0	1	1	0	3	1
18	18	0	5	1	0	1	1	0	4	1
19	19	0	3	1	0	1	1	0	2	1
20	20	14	5	0.02	0.3	1	0.61	13.8	4	0.01
21	21	0	5	1	0	1	1	0	4	1
22	22	0	4	1	0	1	1	0	3	1
23	23	0	5	1	0	1	1	0	4	1
24	24	0	4	1	0	1	1	0	3	1
25	25	0	4	1	0	1	1	0	3	1
26	26	0	4	1	0	1	1	0	3	1
27	27	0	5	1	0	1	1	0	4	1
28	28	0	4	1	0	1	1	0	3	1
29	29	0	5	1	0	1	1	0	4	1
30	30	0	4	1	0	1	1	0	3	1

Figures

Figure 1

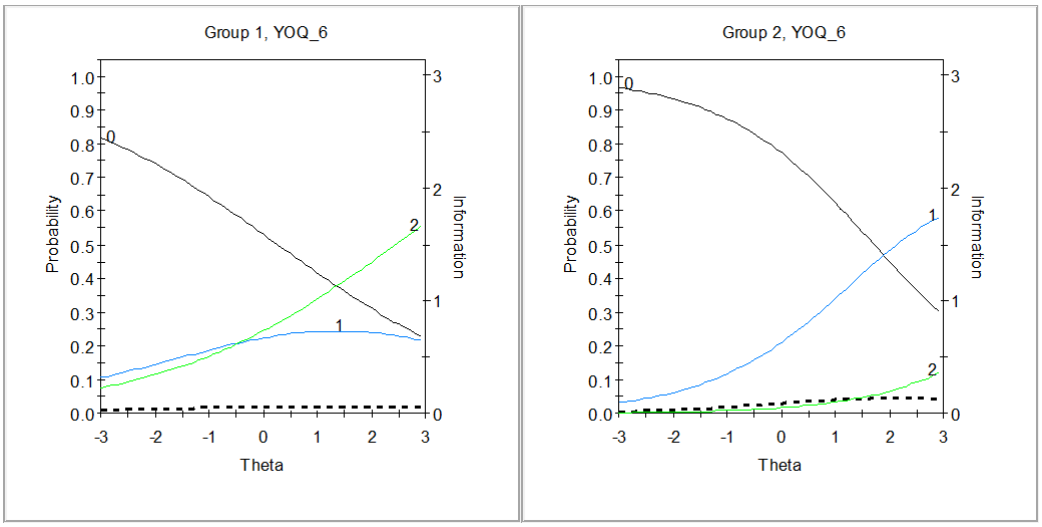
Item Characteristic Curves for all Items showing DIF-Parents vs. Children T2



Note: Group 1=Parents; Group 2=Children

Figure 2

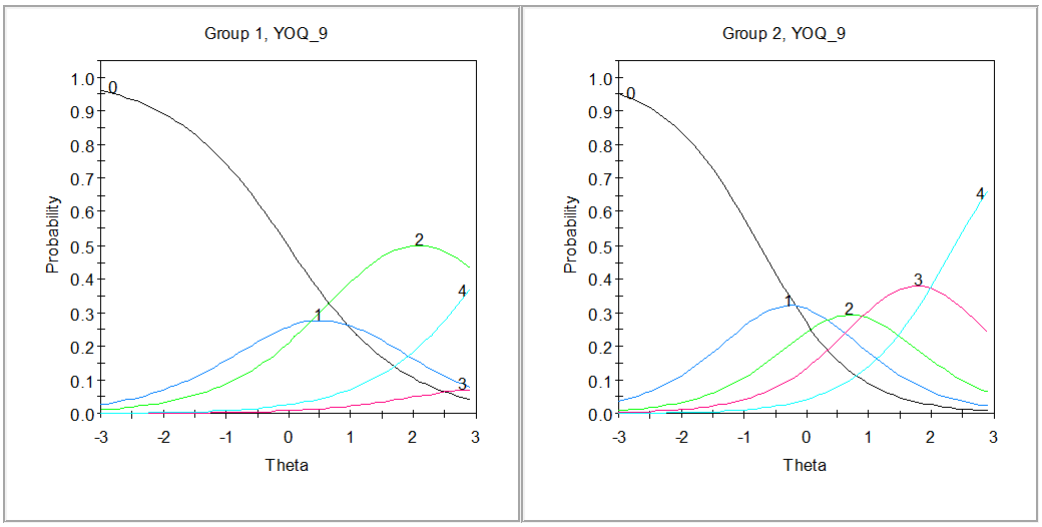
Item Characteristic Curves for all Items showing DIF-Child T1 vs. Child T2

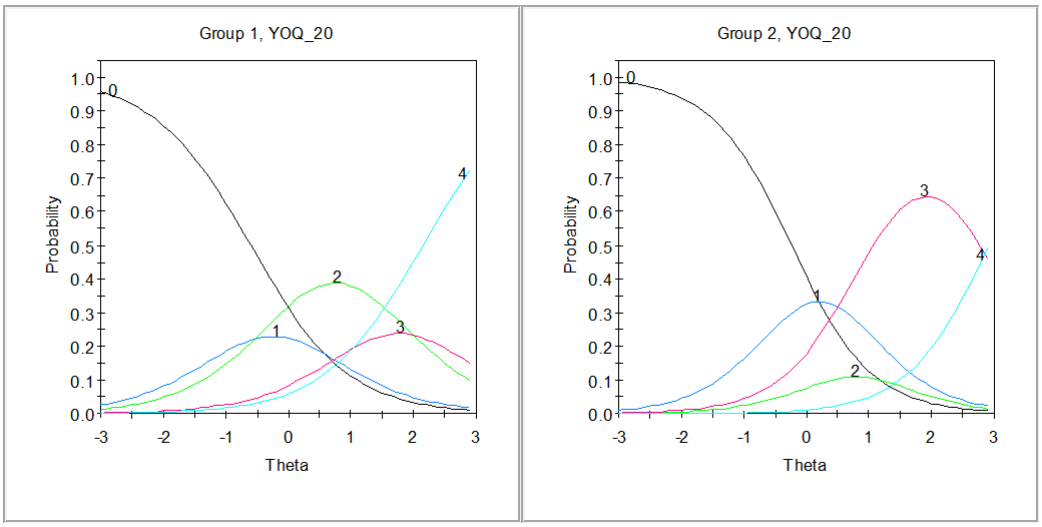


Note: Group 1=T1; Group 2=T2

Figure 3

Item Characteristic Curves for all Items showing DIF-Parent T1 vs. Parent T2





Note: Group 1=T1; Group 2=T2

References

Achenbach, T. M., Mcconaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin,101*(2), 213-232. doi:10.1037//0033-2909.101.2.213

Achenbach, T. M. (1991). *Manual for the Youth Self-Report and 1991 Profile*. Burlington, VT : University of Vermont Department of Psychiatry.

Achenbach, T. M. (2006). As Others See Us. *Current Directions in Psychological Science,15*(2), 94-98. doi:10.1111/j.0963-7214.2006.00414.x

American Psychiatric Publishing. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington (D.C.).

Barmish, A. J., & Kendall, P. C. (2005). Should Parents Be Co-Clients in Cognitive-Behavioral Therapy for Anxious Youth? *Journal of Clinical Child & Adolescent Psychology*, *34*(3), 569–581. doi: 10.1207/s15374424jccp3403_12

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC neuroscience, 11*, 118. https://doi.org/10.1186/1471-2202-11-118

Battagliese, G., Caccetta, M., Luppino, O. I., Baglioni, C., Cardi, V., Mancini, F., & Buonanno, C. (2015). Cognitive-behavioral therapy for externalizing disorders: A meta-analysis of treatment effectiveness. *Behaviour Research and Therapy*, *75*, 60–71. doi: 10.1016/j.brat.2015.10.008

Beck, A. T. (1979). *Cognitive therapy of depression*. New York: Guilford Press.

Biswas, J., Gangadhar, B., & Keshavan, M. (2016). Cross cultural variations in psychiatrists' perception of mental illness: A tool for teaching culture in psychiatry. *Asian Journal of Psychiatry*, *23*, 1–7. doi: 10.1016/j.ajp.2016.05.011

Brown, N. M., Brown, S. N., Briggs, R. D., Germán, M., Belamarich, P. F., & Oyeku, S. O. (2017). Associations Between Adverse Childhood Experiences and ADHD Diagnosis and Severity. *Academic Pediatrics*, *17*(4), 349–355. doi: 10.1016/j.acap.2016.08.013

Budson, A. E., & Price, B. H. (2005). Memory Dysfunction. *New England Journal of Medicine*, *352*(7), 692–699. doi: 10.1056/nejmra041071

Burlingame, G. M., Mosier, J. I., Wells, M. G., Atkin, Q. G., Lambert, M. J., Whoolery, M., et al. (2001). Tracking the influence of mental health treatment: The development of the Youth Outcome Questionnaire. Clinical Psychology and Psychotherapy, 8, 361-379.

Burlingame, G. M., Wells, M. G., Lambert, M. J., & Cox, J. C. (2004). Youth Outcome Questionnaire (Y-OQ). In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Instruments for children and adolescents* (pp. 235-273).

Campbell, N., Ali, F., Finlay, A. Y., & Salek, S. S. (2015). Equivalence of electronic and paper-based patient-reported outcome measures. *Quality of Life Research*, *24*(8), 1949–1961. doi: 10.1007/s11136-015-0937-3

Chorpita, B., & Weisz, J. (2009). *Match-ADTC: Modular Approach to Therapy for Children (Match) with Anxiety, Depression, Trauma, or Conduct Problems*. Satellite Beach, FL: PracticeWise, LLC.

Cleridou, K., Patalay, P., & Martin, P. (2017). Does parent–child agreement vary based on presenting problems? Results from a UK clinical sample. *Child and Adolescent Psychiatry and Mental Health*, *11*(1). doi: 10.1186/s13034-017-0159-2

Cook, I. A., Balasubramani, G., Eng, H., Friedman, E., Young, E. A., Martin, J., et al. (2007). Electronic source materials in clinical research: Acceptability and validity of symptom self-rating in major depressive disorder. *Journal of Psychiatric Research*, *41*(9), 737–743. doi: 10.1016/j.jpsychires.2006.07.015

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring Informant Discrepancies in Clinical Child Research. *Psychological Assessment,16*(3), 330-334. doi:10.1037/1040-3590.16.3.330

De Los Reyes, A., & Kazdin, A. E. (2005). Informant Discrepancies in the Assessment of Childhood Psychopathology: A Critical Review, Theoretical Framework, and Recommendations for Further Study. *Psychological Bulletin,131*(4), 483-509. doi:10.1037/0033-2909.131.4.483

De Los Reyes, A., & Kazdin, A. E. (2006). Conceptualizing changes in behavior in intervention research: The range of possible changes model. *Psychological Review,113*(3), 554-583. doi:10.1037/0033-295x.113.3.554

De Los Reyes, A., Kundey, S. M., & Wang, M. (2011). The end of the primary outcome measure: A research agenda for constructing its replacement. *Clinical Psychology Review,31*(5), 829-838. doi:10.1016/j.cpr.2011.03.011

De Los Reyes, A., Aldao, A., Thomas, S. A., Daruwala, S., Swan, A. J., Wie, M. V., Lechner, W. V. (2012). Adolescent Self-Reports of Social Anxiety: Can They Disagree with Objective Psychophysiological Measures and Still Be

Valid? *Journal of Psychopathology and Behavioral Assessment,34*(3), 308-322. doi:10.1007/s10862-012-9289-2

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin,141*(4), 858-900. doi:10.1037/a0038498

Dirks, M. A., Boyle, M. H., & Georgiades, K. (2011). Psychological Symptoms in Youth and Later Socioeconomic Functioning: Do Associations Vary by Informant? *Journal of Clinical Child & Adolescent Psychology,40*(1), 10-22. doi:10.1080/15374416.2011.533403

Duff, K., Beglinger, L. J., Schultz, S. K., Moser, D. J., McCaffrey, R. J., Haase, R. F., Westervelt, H. J., Langbehn, D. R., Paulsen, J. S., & Huntington's Study Group (2007). Practice effects in the prediction of long-term cognitive outcome in three patient samples: a novel prognostic index. *Archives of clinical neuropsychology: the official journal of the National Academy of Neuropsychologists*, *22*(1), 15–24. https://doi.org/10.1016/j.acn.2006.08.013

Dunn, T. W., Burlingame, G. M., Walbridge, M., Smith, J., & Crum, M. J. (2005). Outcome assessment for children and adolescents: psychometric validation of the Youth Outcome Questionnaire 30.1 (Y-OQ®-30.1). *Clinical Psychology & Psychotherapy,12*(5), 388-401. doi:10.1002/cpp.461

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring Depression Over Time . . . or not? Lack of Unidimensionality and Longitudinal Measurement Invariance in Four Common

Rating Scales of Depression. *Psychological Assessment*. Advance online publication. http://dx.doi.org/10.1037/pas0000275

Goldberg, T. E., Harvey, P. D., Wesnes, K. A., Snyder, P. J., & Schneider, L. S. (2015). Practice effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimer's & dementia (Amsterdam, Netherlands), 1*(1), 103–111. https://doi.org/10.1016/j.dadm.2014.11.003

Goldstein, L. A., Gibbons, M. B. C., Thompson, S. M., Scott, K., Heintz, L., Green, P., Crits-Christoph, P. (2010). Outcome Assessment via Handheld Computer in Community Mental Health: Consumer Satisfaction and Reliability. *The Journal of Behavioral Health Services & Research*, *38*(3), 414–423. doi: 10.1007/s11414-010-9229-4

Goodman, K. L., Reyes, A. D., & Bradshaw, C. P. (2010). Understanding and Using Informants' Reporting Discrepancies of Youth Victimization: A Conceptual Model and Recommendations for Research. *Clinical Child and Family Psychology Review*,*13*(4), 366-383. doi:10.1007/s10567-010-0076-x

Gwaltney, C. J., Shields, A. L., & Shiffman, S. (2008). Equivalence of Electronic and Paper-and-Pencil Administration of Patient-Reported Outcome Measures: A Meta-Analytic Review. *Value in Health*, *11*(2), 322–333. doi: 10.1111/j.1524-4733.2007.00231.x

Handler, Marcie & DuPaul, George. (2005). Assessment of ADHD: Differences Across Psychology Specialty Areas. *Journal of attention disorders, 9:* 402-12. doi:10.1177/1087054705278762.

Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice, 35*, 485-491.

Haque, A. (2010). Mental health concepts in Southeast Asia: Diagnostic considerations and treatment implications. *Psychology, Health & Medicine*, *15*(2), 127–134. doi: 10.1080/13548501003615266

Hawley, K. M., & Weisz, J. R. (2003). Child, parent and therapist (dis)agreement on target problems in outpatient therapy: The therapists dilemma and its implications. *Journal of Consulting & Clinical Psychology,71*(1), 62-70. doi:10.1037//0022-006x.71.1.62

Hayes, L., Harvey, C., & Farhall, J. (2013). Family psychoeducation for the treatment of psychosis. *InPsych*, *35*(2).

Hicks, B. M., Blonigen, D. M., Kramer, M. D., Krueger, R. F., Patrick, C. J., Iacono, W. G., & McGue, M. (2007). Gender differences and developmental change in externalizing disorders from late adolescence to early adulthood: A longitudinal twin study. *Journal of abnormal psychology*, *116*(3), 433–447. https://doi.org/10.1037/0021-843X.116.3.433

Jerusalmi, D. (May, 2015). *Cross-Informant Discrepancy Within Child Psychotherapy in a Longitudinal Sample.* Paper presented at the Annual Meeting of the Association for Psychological Science, New York, NY.

Joubert, T., & Kriek, H. J. (2009). Psychometric comparison of paper-and-pencil and online personality assessments in a selection setting. *SA Journal of Industrial Psychology*, *35*(1). doi: 10.4102/sajip.v35i1.727

Kelly, A. M. C., & Garavan, H. (2004). Human Functional Neuroimaging of Brain Changes Associated with Practice. *Cerebral Cortex, 15*(8), 1089–1102. doi: 10.1093/cercor/bhi005

Kelley, M. L., Bravo, A. J., Hamrick, H. C., Braitman, A. L., White, T. D., & Jenkins, J. (2017). Parents' Reports of Children's Internalizing Symptoms: Associations with Parents' Mental Health Symptoms and Substance Use Disorder. *Journal of child and family studies*, *26*(6), 1646–1654. https://doi.org/10.1007/s10826-017-0677-9

Kendall, P. C. (2000). *Cognitive-behavioral therapy for anxious children—therapist manual* (2nd ed.). Ardmore, PA: Workbook Publishing.

Kingery, J. N., Roblek, T. L., Suveg, C., Grover, R. L., Sherrill, J. T., & Bergman, R. L. (2006). They're Not Just "Little Adults": Developmental Considerations for Implementing Cognitive-Behavioral Therapy With Anxious Youth. *Journal of Cognitive Psychotherapy, 20*(3), 263–273. doi: 10.1891/jcop.20.3.263

Koss & Shiang. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), Handbook of psychotherapy and behavior change (4th ed., pp. 664-700). New York: John Wiley & Sons, Inc.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A New Approach to Integrating Data From Multiple Informants in Psychiatric Assessment and Research: Mixing and Matching Contexts and Perspectives. *American Journal of Psychiatry,160*(9), 1566-1577. doi:10.1176/appi.ajp.160.9.1566

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on

patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11,* 49-68.

Larson, K., Russ, S.A., Kahn, R.S., Halfon, N. (2011). Patterns of comorbidity, functioning, and service use for US children with ADHD, 2007. *Pediatrics, 127*(3), 462-70. doi: 10.1542/peds.2010-0165

Levesque, R. J. (2011). Externalizing and Internalizing Symptoms. *Encyclopedia of Adolescence,*903-905. doi:10.1007/978-1-4419-1695-2_539

Lochman, J. E., Powell, N. P., Boxmeyer, C. L., & Jimenez-Camargo, L. (2011). Cognitive-Behavioral Therapy for Externalizing Disorders in Children and Adolescents. *Child and Adolescent Psychiatric Clinics of North America*, *20*(2), 305–318. doi: 10.1016/j.chc.2011.01.005

Lord, F. M. (1958). The utilization of unreliable difference scores. *Journal of Educational Psychology*, *49*(3), 150–152. doi: 10.1037/h0045456

Lukens, E. P., & Mcfarlane, W. R. (2004). Psychoeducation as Evidence-Based Practice: Considerations for Practice, Research, and Policy. *Brief Treatment and Crisis Intervention*, *4*(3), 205–225. doi: 10.1093/brief-treatment/mhh019

McClendon, Debra. (2009). Relative Sensitivity to Change of Psychotherapy Outcome Measures for Children and Adolescents: A Comparison Using Parent- and Self-Report Versions of the CBCL/6-18, BASC-2, and Y-OQ-2.01. *All Theses and Dissertations.* 2089. https://scholarsarchive.byu.edu/etd/2089

Owens, J. S., & Hoza, B. (2003). The role of inattention and hyperactivity/impulsivity in the positive illusory bias. *Journal of Consulting and Clinical Psychology*, *71*(4), 680–691. doi: 10.1037/0022-006x.71.4.680

Oud, M., Winter, L. D., Vermeulen-Smit, E., Bodden, D., Nauta, M., Stone, L., van den Heuvel, M., Taher, RA., de Graaf, I., Kendall, T., Engles, R., & Stikkelbroek, Y. (2019). Effectiveness of CBT for children and adolescents with depression: A systematic review and meta-regression analysis. *European Psychiatry*, *57*, 33–45. doi: 10.1016/j.eurpsy.2018.12.008

Peper, J. S., & Dahl, R. E. (2013). Surging Hormones: Brain-Behavior Interactions During Puberty. *Current directions in psychological science*, *22*(2), 134–139. doi.org/10.1177/0963721412473755

Popp, L., Neuschwander, M., Mannstadt, S., In-Albon, T., & Schneider, S. (2017). Parent-Child Diagnostic Agreement on Anxiety Symptoms with a Structured Diagnostic Interview for Mental Disorders in Children. *Frontiers in Psychology*, *8*. doi: 10.3389/fpsyg.2017.00404

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin,114*(3), 552-566. doi:10.1037//0033-2909.114.3.552

Sanchez, L. M., & Turner, S. M. (2003). Practicing psychology in the era of managed care: Implications for practice and training. *American Psychologist, 58*(2), 116-129.

Sauter, F. M., Heyne, D., & Michiel Westenberg, P. (2009). Cognitive behavior therapy for anxious adolescents: developmental influences on treatment design and delivery. *Clinical child and family psychology review*, *12*(4), 310–335. https://doi.org/10.1007/s10567-009-0058-z

Stallard, P. (2009). *Anxiety: Cognitive behavior therapy with children and young people*. East Sussex, UK: Routledge

Stikkelbroek, Y., Bodden, D. H., Deković, M., & van Baar, A. L. (2013). Effectiveness and cost effectiveness of cognitive behavioral therapy (CBT) in clinically depressed adolescents: individual CBT versus treatment as usual (TAU). *BMC psychiatry*, *13*, 314. https://doi.org/10.1186/1471-244X-13-314

Tay, L., Meade, A. W., & Cao, M. (in press). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*. doi: 10.1177/1094428114553062

Van de Looij-Jansen, P. M., Jansen, W., Wilde, E. J., Donker, M. C., & Verhulst, F. C. (2010). Discrepancies Between Parent-Child Reports of Internalizing Problems Among Preadolescent Children: Relationships with Gender, Ethnic Background, and Future Internalizing Problems. *The Journal of Early Adolescence,31*(3), 443-462. doi:10.1177/0272431610366243

Van Droogenbroeck, F., Spruyt, B., & Keppens, G. (2018). Gender differences in mental health problems among adolescents and the role of social support: results from the Belgian health interview surveys 2008 and 2013. *BMC psychiatry*, *18*(1), 6. https://doi.org/10.1186/s12888-018-1591-4

Weisz, J. R., & Mccarty, C. A. (1999). Can we trust parent reports in research on cultural and ethnic differences in child psychopathology? Using the bicultural family design to test parental culture effects. *Journal of Abnormal Psychology,108*(4), 598-605. doi:10.1037//0021-843x.108.4.598

Weisz, J., Hawley, K., & Jenson Doss, A. (2004). Empirically tested psychotherapies for youth internalizing and externalizing problems and disorders. *Child and Adolescent Psychiatric Clinics of North America*, *13*(4), 729–815. doi: 10.1016/j.chc.2004.05.006

Wells, M. G., Burlingame, G. M., Lambert, M. J., Hoag, M. J., & Hope, C. A. (1996). Conceptualization and measurement of patient change during psychotherapy: Development of the Outcome Questionnaire and Youth Outcome Questionnaire. *Psychotherapy*, *33*, 275-283.

Wells, M.G., Burlingame, G.M., & Rose, P.M. (2003). Youth outcome questionnaire self report. Wilmington, DE: American Professional Credentialing Services.

Winarick, J. (2018). Lack of Measurement Invariance Between Parent and Child Report on the Youth Outcome Questionnaire. Manuscript in preparation.

Wolford, G., Rosenberg, S. D., Rosenberg, H. J., Swartz, M. S., Butterfield, M. I., Swanson, J. W., & Jankowski, M. K. (2008). A Clinical Trial Comparing Interviewer and Computer-Assisted Assessment Among Clients With Severe Mental Illness. *Psychiatric Services*, *59*(7), 769–775. doi: 10.1176/ps.2008.59.7.769

Vita

Name	*Jenna Winarick*
Baccalaureate Degree	*Bachelor of Arts, University of Maryland, College Park, MD, Major: Psychology and English Literature*
Date Graduated	*May, 2012*
Other Degrees and Certificates	*Master of Arts, Teachers College, Columbia University, New York, NY, Major: Clinical Psychology*
Date Graduated	*May, 2015*
Other Degrees and Certificates	*Master of Arts, St. John's University, Queens, NY, Major: Clinical Psychology*
Date Graduated	*May, 2018*